

CURRENT STATUS LINEAR REGRESSION

BY PIET GROENEBOOM AND KIM HENDRICKX

Delft University of Technology and Hasselt University

We construct \sqrt{n} -consistent and asymptotically normal estimates for the finite dimensional regression parameter in the current status linear regression model, which do not require any smoothing device and are based on maximum likelihood estimates (MLEs) of the infinite dimensional parameter. We also construct estimates, again only based on these MLEs, which are arbitrarily close to efficient estimates, if the generalized Fisher information is finite. This type of efficiency is also derived under minimal conditions for estimates based on smooth non-monotone plug-in estimates of the distribution function. Algorithms for computing the estimates and for selecting the bandwidth of the smooth estimates with a bootstrap method are provided. The connection with results in the econometric literature is also pointed out.

1. Introduction. Investigating the relationship between a response variable Y and one or more explanatory variables is a key activity in statistics. Often encountered in regression analysis however, are situations where a part of the data is not completely observed due to some kind of censoring. In this paper we focus on modeling a linear relationship when the response variable is subject to interval censoring type I, i.e. instead of observing the response Y , one only observes whether or not $Y \leq T$ for some random censoring variable T , independent of Y . This type of censoring is often referred to as the current status model and arises naturally, for example, in animal tumorigenicity experiments (see e.g. [8] and [9]) and in HIV and AIDS studies (see e.g. [32]). Substantial literature has been devoted to regression models with current status data including the proportional hazard model studied in [19], the accelerated failure time model proposed by [26] and the proportional odds regression model of [27].

The regression model we want to study is the semi-parametric linear regression model $Y = \beta'_0 X + \varepsilon$, where the error terms are assumed to be independent of T and X with unknown distribution function F_0 . This model is closely related to the binary choice model type, studied in econometrics (see e.g. [3, 5], [21] and [7]), where, however, the censoring variable T is degenerate, i.e. $P(T = 0) = 1$, and observations are of the type $(X_i, 1_{\{Y_i \leq 0\}})$. In the latter model, the scale is not identifiable, which

AMS 2000 subject classifications: Primary 62G05, 62N01; secondary 62-04

Keywords and phrases: current status, linear regression, MLE, semi-parametric model

one usually solves by adding a constraint on the parameter space such as setting the length of β or the first coefficient equal to one.

Our model of interest is parametrized by the finite dimensional regression parameter β_0 and the infinite dimensional nuisance parameter F_0 that contains β_0 as one of its arguments. A similar bundled parameter problem was studied by [6], where the authors first provide a framework for the distributional theory of problems with bundled parameters and next prove their theory for efficient estimation in the linear regression model with right censored data. A spline based estimate of the nuisance parameter is proposed.

Although it is indeed tempting to think that some kind of smoothing is needed, like the splines in [6] or the kernel estimates in the econometric literature for the binary choice model, where even higher order kernels are used (see, e.g., [21]), a maximum rank correlation estimate which does not use any smoothing has been introduced in [16], and this estimator has been proved to be \sqrt{n} -consistent and asymptotically normal in [31]. However, the latter estimate does not attain the efficiency bounds and one wonders whether it is possible to construct simple discrete estimates of this type and achieve the efficiency bounds. It is not clear how the maximum rank correlation estimate in [16] could be used to this end, and we therefore turn to estimators depending on maximum likelihood estimators for the nuisance parameter.

The profile maximum likelihood estimator (MLE) of β_0 was proved to be consistent in [3] but nothing is known about its asymptotic distribution, apart from its consistency and upper bounds for its rate of convergence. It remains an open question whether or not the profile MLE of β_0 is \sqrt{n} -consistent. [24] derived an $n^{1/3}$ -rate for the profile MLE; we show that without any smoothing it is possible to construct estimates, based on the MLE for the distribution function F for fixed β , that converge at \sqrt{n} -rate to the true parameter. We note, however, that the estimator we propose, based on the nonparametric MLE for F for fixed β , is *not* the profile MLE for β_0 . The estimator is a kind of hybrid estimator, which is based on the argmax MLE for F for fixed β , but defined as the zero of a non-smooth score function as a function of β . So we have the remarkable situation that finding the estimate $\hat{\beta}_n$ as the root of a score equation based on the MLEs $\hat{F}_{n,\beta}$, can be proved to give \sqrt{n} -consistent estimates of β_0 , in contrast with the argmax approach, using profile likelihood, for which we even still do not know whether it is \sqrt{n} -consistent. We go somewhat deeper into this matter in the discussion section of this paper.

A general theoretical framework for semi-parametric models when the criterion function is not smooth is developed in [2]. The proposed theory is less suited for our score approach since the authors assume existence of a uniform consistent estimator for the infinite dimensional regression parameter with convergence rate not depending on the finite dimensional regression parameter of interest. In the current status linear regression model we have to estimate β_0 and F_0 simultaneously, as a consequence

the convergence rate of the estimator for F_0 depends on the convergence rate of the estimator for β_0 , the parameters β_0 and F_0 are bundled and therefore we cannot apply their theory.

[24] considers efficient estimation for the current status model with a 1-dimensional regression parameter β via a penalized maximum likelihood estimator under the conditions that F_0 and $u \mapsto E_\beta(X|T - \beta X = u)$ are three times continuously differentiable and that the data only provide information about a part of the distribution function F_0 , where F_0 stays away from zero and 1. [23] proposes an estimation equation for β , derived from an inequality on the conditional covariance between X and Δ conditional on $T - \beta'X$, and use a U-statistics representation, involving summation over many indices. [30] considered an estimator based on a random sieved likelihood, but the expression for the efficient information (based on the generalized Fisher information) in this paper seems to be different from what we and the authors mentioned above obtain for this expression.

Approaches to \sqrt{n} -consistent and efficient estimation of the regression parameters in the binary choice model were considered by [21] and [5] among others. For a derivation of the efficient information $\tilde{\ell}_{\beta_0, F_0}^2$, defined by

$$(1.1) \quad \begin{aligned} \tilde{\ell}_{\beta, F}(t, x, \delta) = & \left\{ E(X|T - \beta'X = t - \beta'x) - x \right\} f(t - \beta'x) \\ & \cdot \left\{ \frac{\delta}{F(t - \beta'x)} - \frac{1 - \delta}{1 - F(t - \beta'x)} \right\}, \end{aligned}$$

where we assume $f(t - \beta'x) > 0$, we refer to [4] for the binary choice model, and next to [20] and [24] for the current status regression model.

As mentioned above, the condition that the support of the density of $T - \beta'_0 X$ is strictly contained in an interval D for all β and that F_0 stays strictly away from 0 and 1 on D is used in [24]. This condition is also used in [20] and [30]. The drawback of the assumption is that we have no information about the whole distribution F_0 . It also goes against the usual conditions made for the current status model, where one commonly assumes that the observations provide information over the whole range of the distribution one wants to estimate. We presume that this assumption is made for getting the Donsker properties to work and to avoid truncation devices that can prevent the problems arising if this condition is not made, such as unbounded score functions and ensuing numerical difficulties. Examples of truncation methods can be found in [5] and [21] among others where the authors consider truncation sequences that converge to zero with increasing sample size. We show that it is possible to estimate the finite dimensional regression parameter β_0 at \sqrt{n} -rate based on a fixed truncated subsample of the data where the truncation area is determined by the quantiles of the infinite dimensional nuisance parameter estimator.

The paper is organized as follows. The model, its corresponding log likelihood and a truncated version of the log likelihood are introduced in Section 2. In this section,

we also discuss the advantages of a score approach over the maximum likelihood characterization. The behavior of the MLE for the distribution function F_0 in case β is not equal to β_0 is studied in Section 3. We first construct in Section 4, based on a score equation, a \sqrt{n} -consistent but inefficient estimate of the regression parameter based on the MLE of F_0 and show how an estimate of the density, based on the MLE, can be used to extend the estimate of the regression parameter to an estimate with an asymptotic variance that is arbitrarily close to the information lower bound.

Next, we give the asymptotic behavior of a plug-in estimator which is obtained by a score equation derived from the truncated log likelihood in case a second order kernel estimate for the distribution function F_0 is considered. We show that the latter estimator is \sqrt{n} -consistent and asymptotically normal with an asymptotic variance that is arbitrarily (determined by the truncation device) close to the information lower bound, just like the estimator based on the MLE we discussed in the preceding paragraph.

The estimation of an intercept term, that originates from the mean of the error distribution, is outlined in Section 5. Section 6 contains details on the computation of the estimates together with the results of our simulation study; a bootstrap method for selecting a bandwidth parameter is also given. A discussion of our results is given in Section 7. Section 8 contains the derivation of the efficient information given in (1.1). The proofs of the results given in this paper are worked out in the supplemental article [12].

2. Model description. Let $(T_i, X_i, \Delta_i), i = 1, \dots, n$ be independent and identically distributed observations from $(T, X, \Delta) = (T, X, 1_{\{Y \leq T\}})$. We assume that Y is modeled as

$$(2.1) \quad Y = \beta_0' X + \varepsilon,$$

where β_0 is a k -dimensional regression parameter in the parameter space Θ and ε is an unobserved random error, independent of (T, X) with unknown distribution function F_0 . We assume that the distribution of (T, X) does not depend on (β_0, F_0) which implies that the relevant part of the log likelihood for estimating (β_0, F_0) is given by:

$$(2.2) \quad \begin{aligned} l_n(\beta, F) &= \sum_{i=1}^n [\Delta_i \log F(T_i - \beta' X_i) + (1 - \Delta_i) \log \{1 - F(T_i - \beta' X_i)\}] \\ &= \int [\delta \log F(t - \beta' x) + (1 - \delta) \log \{1 - F(t - \beta' x)\}] d\mathbb{P}_n(t, x, \delta), \end{aligned}$$

where \mathbb{P}_n is the empirical distribution of the (T_i, X_i, Δ_i) . We will denote the probability measure of (T, X, Δ) by P_0 . We define the truncated log likelihood $l_n^{(\epsilon)}(\beta, F)$

by

$$(2.3) \quad \int_{F(t-\beta'x) \in [\epsilon, 1-\epsilon]} [\delta \log F(t-\beta'x) + (1-\delta) \log \{1 - F(t-\beta'x)\}] d\mathbb{P}_n(t, x, \delta),$$

where $\epsilon \in (0, 1/2)$ is a truncation parameter. Analogously, let,

$$(2.4) \quad \psi_n^{(\epsilon)}(\beta, F) = \int_{F(t-\beta'x) \in [\epsilon, 1-\epsilon]} \phi(t, x, \delta) \{F(t-\beta'x) - \delta\} d\mathbb{P}_n(t, x, \delta),$$

define the truncated score function for some weight function ϕ . In this paper, we consider estimates of β_0 , obtained by solving a score equation $\psi_n^{(\epsilon)}(\beta, \hat{F}_\beta) = 0$ where \hat{F}_β is an estimate of F for fixed β . A motivation of the score approach is outlined below, we have three reasons for using the score function characterization instead of the argmax approach for the estimation of β_0 .

- (i) Our simulation experiments indicate that, even if the profile MLE would be \sqrt{n} -consistent, its variance is clearly bigger than the other estimates we propose.
- (ii) The characterization of $\hat{\beta}_n$ as the solution of a score equation

$$\psi_n(\beta, \hat{F}_{n,\beta}) = 0,$$

(see, e.g., (2.4)), where $\hat{F}_{n,\beta}$ is the MLE for fixed β maximizing the log likelihood defined in (2.2) over all $F \in \mathcal{F} = \{F : \mathbb{R} \rightarrow [0, 1] : F \text{ is a distribution function}\}$, gives us freedom in choosing the function ψ_n of which we try to find the root $\hat{\beta}_n$. Smoothing techniques can be used but are not necessary to obtain \sqrt{n} -convergence of the estimate.

In this paper, we first choose a function ψ_n , which produces a \sqrt{n} -consistent and asymptotically normal estimate of β_0 , and does not need any smoothing device. Just like the Han maximum correlation estimate, this estimate does not attain the efficiency bound, although the difference between its asymptotic variance and the efficient asymptotic variance is rather small in our experiments. More details are given in Section 6.

Next we choose a function ψ_n which gives (only depending on our truncation device) an asymptotic variance which is arbitrarily close to the efficient asymptotic variance. In this case we need an estimate of the density of the error distribution and are forced to use smoothing in the definition of ψ_n . The estimate, although efficient in the sense we use this concept in our paper, is not necessarily better in small samples, though.

- (iii) The “canonical” approach to proofs that argmax estimates of β_0 are \sqrt{n} -consistent has been provided by [31]. His Theorem 1 says that $\|\hat{\beta}_n - \beta_0\| = O_p(n^{-1/2})$, where $\|\cdot\|$ denotes the Euclidean norm, if $\hat{\beta}_n$ is the maximizer of $\Gamma_n(\beta)$, with population equivalent $\Gamma(\beta)$ and

(a) there exists a neighborhood N of β_0 and a constant $k > 0$ such that

$$\Gamma(\beta) - \Gamma(\beta_0) \leq -k\|\beta - \beta_0\|^2,$$

for $\beta \in N$, and

(b) uniformly over $o_p(1)$ neighborhoods of β_0 ,

$$\begin{aligned} \Gamma_n(\beta) - \Gamma_n(\beta_0) \\ = \Gamma(\beta) - \Gamma(\beta_0) + O_p(\|\beta - \beta_0\|/\sqrt{n}) + o_p(\|\beta - \beta_0\|^2) + O_p(n^{-1}). \end{aligned}$$

If we try to apply this to the profile MLE $\hat{\beta}_n$, it is not clear that an expansion of this type will hold. We seem to get inevitably an extra term of order $O_p(n^{-2/3})$ in (b), which does not fit into this framework. On the other hand, in the expansion of our score function ψ_n , we get that this function is in first order the sum of a term of the form

$$\psi'(\beta_0)(\beta - \beta_0)$$

where ψ' is the matrix, representing the total derivative of the population equivalent score function ψ , and a term W_n of order $O_p(n^{-1/2})$, which gives:

$$\hat{\beta}_n - \beta_0 \sim -\psi'(\beta_0)^{-1}W_n = O_p(n^{-1/2}),$$

and here extra terms of order $O_p(n^{-2/3})$ do not hurt. The technical details are elaborated in the proofs of our main result given in the supplemental article [12].

Before we formulate our estimates, we first describe in Section 3 the behavior of the MLE $\hat{F}_{n,\beta}$ for fixed β . Throughout the paper, we illustrate our estimates by a simple simulated data example; we consider the model $Y_i = 0.5X_i + \varepsilon_i$, where the X_i and T_i are independent $\text{Uniform}(0, 2)$ and where the ε_i are independent random variables with density $f(u) = 384(u - 0.375)(0.625 - u)1_{[0.375, 0.625]}(u)$ and independent of the X_i and T_i . Note that the expectation of the random error $E(\varepsilon) = 0.5$, our linear model contains an intercept, $\mathbb{E}(Y_i|X_i = x_i) = 0.5 + 0.5x_i$.

REMARK 2.1. We chose the present model as a simple example of a model for which the (generalized) Fisher information is finite. This Fisher information easily gets infinite. For example, if F_0 is the uniform distribution on $[0, 1]$ and X and T (independently) also have uniform distributions on $[0, 1]$ and $\beta = 1/2$, the Fisher information for estimating β is given by:

$$\int_{u=0}^{1/2} \frac{(x - 1/2)^2}{u(1 - u)} dx du + \int_{u=1/2}^1 \frac{\{x - (1 - u)\}^2}{u(1 - u)} dx du = \infty.$$

We observed in simulations with the uniform distribution that n times the variance of our estimates (using $\epsilon = 0$) steadily decreases with increasing sample size n , suggesting a faster than \sqrt{n} -convergence for the estimate in this model. The theoretical framework for estimation of models with infinite Fisher information falls beyond the scope of this paper. So we chose a model where the ratio $f_0(x)^2/\{F_0(x)\{1 - F_0(x)\}\}$ stays bounded near the boundary of its support by taking a rescaled version of the density $6x(1 - x)1_{[0,1]}(x)$ for f_0 . Note that, if the Fisher information is infinite, our theory still makes sense for the truncated version:

$$\int_{F_0(u) \in [\epsilon, 1-\epsilon]} \int_{x=0}^1 \frac{(x - \mathbb{E}\{X|T - X/2 = u\})^2 f_0(u)^2}{F_0(u)\{1 - F_0(u)\}} f_{X|T-X/2}(x|u) f_{T-X/2}(u) dx du,$$

corresponding to our truncation of the log likelihood and the score function in the sequel. For completeness we included the derivation of the Fisher information in the Appendix. These calculations provide more insight in the information loss when one moves from a parametric model where F_0 is known to our semi-parametric model with unknown F_0 .

3. Behavior of the maximum likelihood estimator. For fixed β , the MLE $\hat{F}_{n,\beta}$ of $l_n(\beta, F_\beta)$ is a piecewise constant function with jumps at a subset of $\{T_i - \beta'X_i : i = 1, \dots, n\}$. Once we have fixed the parameter β , the order statistics on which the MLE is based are the order statistics of the values $U_1^{(\beta)} = T_1 - \beta'X_1, \dots, U_n^{(\beta)} = T_n - \beta'X_n$ and the values of the corresponding $\Delta_i^{(\beta)}$. The MLE can be characterized as the left derivative of the convex minorant of a cumulative sum diagram consisting of the points $(0, 0)$ and

$$(3.1) \quad \left(i, \sum_{j=1}^i \Delta_{(j)}^{(\beta)} \right) \quad i = 1, \dots, n,$$

where $\Delta_{(i)}^{(\beta)}$ corresponds to the i th order statistic of the $T_i - \beta'X_i$ (see e.g. Proposition 1.2 in [15] on p. 41). We have:

$$\mathbb{P}\left\{\Delta_i^{(\beta)} = 1 \mid U_i^{(\beta)} = u\right\} = \int F_0(u + (\beta - \beta_0)'x) f_{X|T-\beta'X}(x|T - \beta'X = u) dx.$$

Hence, defining

$$(3.2) \quad F_\beta(u) = \int F_0(u + (\beta - \beta_0)'x) f_{X|T-\beta'X}(x|T - \beta'X = u) dx,$$

we can consider the $\Delta_i^{(\beta)}$ as coming from a sample in the ordinary current status model, where the observations are of the form $(U_i^{(\beta)}, \Delta_i^{(\beta)})$, and where the observation times have density $f_{T-\beta'X}$ and where $\Delta_i^{(\beta)} = 1$ with probability $F_\beta(U_i^{(\beta)})$ at observation $U_i^{(\beta)}$.

REMARK 3.1. Note that

$$\begin{aligned} F'_\beta(u) &= \int f_0(u + (\beta - \beta_0)'x) f_{X|T-\beta'X}(x|u) dx \\ &\quad + \int F_0(u + (\beta - \beta_0)'x) \frac{\partial}{\partial u} f_{X|T-\beta'X}(x|u) dx. \end{aligned}$$

Integration by parts on the second term yields

$$\begin{aligned} \int F_0(u + (\beta - \beta_0)'x) \frac{\partial}{\partial u} f_{X|T-\beta'X}(x|u) dx \\ = -(\beta - \beta_0)' \int f_0(u + (\beta - \beta_0)'x) \frac{\partial}{\partial u} F_{X|T-\beta'X}(x|u) dx. \end{aligned}$$

This implies

$$\begin{aligned} F'_\beta(u) &= \int f_0(u + (\beta - \beta_0)'x) \left\{ f_{X|T-\beta'X}(x|u) \right. \\ &\quad \left. - (\beta - \beta_0)' \frac{\partial}{\partial u} F_{X|T-\beta'X}(x|u) \right\} dx. \end{aligned}$$

Assuming that $u \mapsto f_{X|T-\beta'X}(x|u)$ stays away from zero on the support of f_0 , this implies by a continuity argument that F_β is monotone increasing on the support of F'_β for β close to β_0 .

Also note that we get from the fact that F_0 is a distribution function with compact support:

$$\lim_{u \rightarrow -\infty} F_\beta(u) = 0 \quad \text{and} \quad \lim_{u \rightarrow \infty} F_\beta(u) = 1.$$

So we may assume that F_β is a distribution function for β close to β_0 .

A picture of the MLE $\hat{F}_{n,\beta}$, based on the values $T_i - \beta X_i$, and the corresponding function F_β for the model used in our simulation experiment, is shown in Figure 1 and compared with F_0 . Note that F_β involves both a location shift and a change in shape of F_0 .

For fixed β in a neighborhood of β_0 we can now use standard theory for the MLE from current status theory. The following assumptions are used.

- (A1) The parameter $\beta_0 = (\beta_{0,1}, \dots, \beta_{0,k}) \in \mathbb{R}^k$ is an interior point of Θ and the parameter space Θ is compact.
- (A2) F_β has a strictly positive continuous derivative, which stays away from zero on $A_{\epsilon',\beta} \stackrel{\text{def}}{=} \{u : F_\beta(u) \in [\epsilon', 1 - \epsilon']\}$ for all $\beta \in \Theta$, where $\epsilon' \in (0, \epsilon)$.
- (A3) The density $u \mapsto f_{T-\beta'X}(u)$ is continuous and also staying away from zero on $A_{\epsilon',\beta}$ for all $\beta \in \Theta$, where $A_{\epsilon',\beta}$ is defined as in (A2).

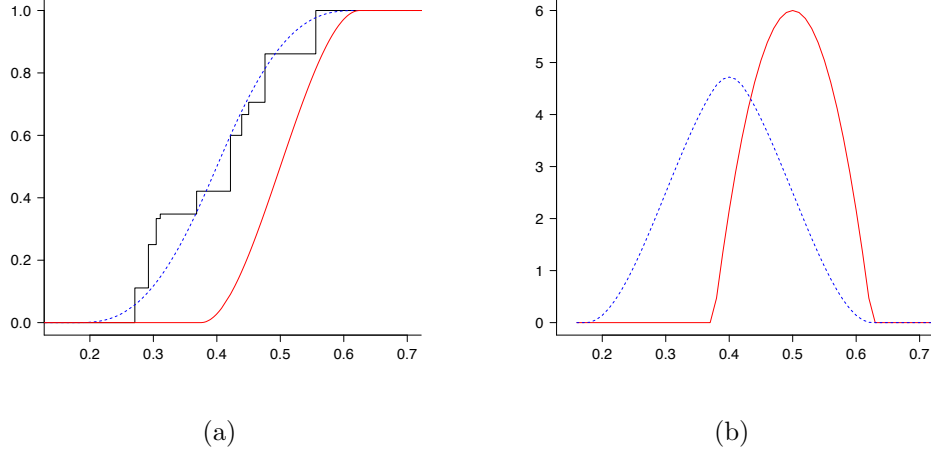


Fig 1: The real F_0 (red, solid), the function F_β for $\beta = 0.6$ (blue, dashed) and the MLE $\hat{F}_{n,\beta}$ (step function), for a sample of size $n = 1000$. (b) The real f_0 (red, solid) and the function F'_β for $\beta = 0.6$ (blue, dashed).

REMARK 3.2. Note that the truncation is for the interval $[\epsilon, 1 - \epsilon]$, but that we need conditions (A2) and (A3) to be satisfied for the slightly bigger interval $[\epsilon', 1 - \epsilon']$.

LEMMA 3.1. *If Assumptions (A1), (A2) and (A3) hold, then:*

(i)

$$\mathbb{P} \left(\lim_{n \rightarrow \infty} \sup_{u \in A_{\epsilon,\beta}} |\hat{F}_{n,\beta}(u) - F_\beta(u)| = 0 \right) = 1$$

(ii)

$$\int_{A_{\epsilon,\beta}} \left\{ \hat{F}_{n,\beta}(u) - F_\beta(u) \right\}^2 du = O_p \left(n^{-2/3} \right).$$

PROOF. (i) is proved in Section 4.1 of Part (ii) in [15] and (ii) follows from (11.34) on p. 327 of [14]. \square

We first show in Section 4.1 that it is possible to construct \sqrt{n} -consistent estimates of β_0 based on solving a score equation $\psi_n^{(\epsilon)}(\beta, \hat{F}_{n,\beta}) = 0$ without requiring any smoothing in the estimation process. In Section 4.2 we look at \sqrt{n} -consistent and efficient score-estimates based on the MLE $\hat{F}_{n,\beta}$, using a weight function ϕ that incorporates the estimate $\int K_h(u - y) d\hat{F}_{n,\beta}(y)$ of the density $f_0(u) = F'_0(u)$. An

efficient estimate of β_0 derived by a score function based on kernel estimates for the distribution function, is considered in Section 4.3. The latter estimate does not involve the behavior of the MLE $\hat{F}_{n,\beta}$.

4. \sqrt{n} -consistent estimation of the regression parameter.

4.1. *A simple estimate based on the MLE $\hat{F}_{n,\beta}$, avoiding any smoothing.* We consider the function $\psi_{1,n}^{(\epsilon)}$, defined by:

$$(4.1) \quad \psi_{1,n}^{(\epsilon)}(\beta) \stackrel{\text{def}}{=} \int_{\hat{F}_{n,\beta}(t-\beta'x) \in [\epsilon, 1-\epsilon]} x \{ \hat{F}_{n,\beta}(t-\beta'x) - \delta \} d\mathbb{P}_n(t, x, \delta),$$

where $\hat{F}_{n,\beta}$ is the MLE based on the order statistics of the values $T_i - \beta'X_i$, $i = 1, \dots, n$ and define the estimate $\hat{\beta}_n$ by,

$$(4.2) \quad \psi_{1,n}^{(\epsilon)}(\hat{\beta}_n) = 0,$$

where 0 is the k -dimensional vector with zeros as components. This “score equation” is inspired by noting that,

$$\begin{aligned} (\beta - \beta_0)' \text{Cov}(\Delta, X | T - \beta'X) &= \text{Cov}(\Delta, (\beta - \beta_0)'X | T - \beta'X) \\ &= \text{Cov}(F_0(T - \beta'X + (\beta - \beta_0)'X), (\beta - \beta_0)'X | T - \beta'X) \geq 0, \end{aligned}$$

for all β , following from the fact that F_0 is an increasing function and $\mathbb{E}(\Delta | T, X) = F_0(T - \beta'X + (\beta - \beta_0)'X)$.

The following assumptions are also needed for the asymptotic normality results of our estimators.

- (A4) The function F_β is twice continuously differentiable on the interior of the support S_β of $f_\beta = F'_\beta$ for $\beta \in \Theta$.
- (A5) The density $f_{T-\beta'X}(u)$ of $T - \beta'X$ and the conditional density $f_{X|T-\beta'X}(x|u)$ of X given $T - \beta'X = u$ are twice continuously differentiable functions w.r.t. u , except possibly at a finite number of points and $\beta \mapsto f_{T-\beta'X}(v)$ and $\beta \mapsto f_{X|T-\beta'X}(x|T - \beta'X = v)$ are continuous functions, for v and x in the definition domain of the functions and for $\beta \in \Theta$. The density of (T, X) has compact support.

THEOREM 4.1. *Let Assumptions (A1)-(A5) be satisfied and suppose that the covariance $\text{covar}(X, F_0(u + (\beta - \beta_0)'X) | T - \beta'X = u)$ is not identically zero for u in the region $A_{\epsilon,\beta}$, for each $\beta \in \Theta$. Then as $n \rightarrow \infty$, we have, $\forall \eta > 0$, $\exists n_0 \in \mathbb{N}$ such that $\forall n > n_0$:*

$$\mathbb{P} \left(\exists \beta \in \{b \in \Theta : \|b - \beta_0\| < \eta\} : \psi_{1,n}^{(\epsilon)}(\beta) = 0 \right) > 1 - \eta.$$

Furthermore, we have, $\hat{\beta}_n \xrightarrow{P} \beta_0$ and $\sqrt{n}\{\hat{\beta}_n - \beta_0\}$ is asymptotically normal with mean zero and variance $A^{-1}BA^{-1}$, where $\hat{\beta}_n$ is defined by (4.2) and

$$A = \mathbb{E}_\epsilon \left[f_0(T - \beta'_0 X) \{X - \mathbb{E}(X|T - \beta'_0 X)\}' \{X - \mathbb{E}(X|T - \beta'_0 X)\} \right],$$

$$B = \mathbb{E}_\epsilon \left[F_0(T - \beta'_0 X) \{1 - F_0(T - \beta'_0 X)\} \{X - \mathbb{E}(X|T - \beta'_0 X)\}' \{X - \mathbb{E}(X|T - \beta'_0 X)\} \right],$$

$\mathbb{E}_\epsilon(w(T, X, \Delta)) = \int_{F_0(t - \beta'_0 x) \in [\epsilon, 1 - \epsilon]} w(t, x, \delta) dP_0(t, x, \delta)$ is the truncated expectation of $w(T, X, \Delta)$ for some deterministic function w and where we also assume that A is non-singular.

REMARK 4.1. Note that $\text{covar}(X, F_0(u + (\beta - \beta_0)'X)|T - \beta'_0 X = u)$ is not identically zero for u in the region $\{u : \epsilon \leq F_\beta(u) \leq 1 - \epsilon\}$ if the conditional distribution of X , given $T - \beta'_0 X = u$, is non-degenerate for some u in this region if F_0 is strictly increasing on $\{u : \epsilon \leq F_\beta(u) \leq 1 - \epsilon\}$.

The proof of Theorem 4.1 is given in the supplemental article [12]. A picture of the truncated profile log likelihood $l_n^{(\epsilon)}(\beta, \hat{F}_{n,\beta})$ and the score function $\psi_{1,n}^{(\epsilon)}(\beta)$ for β ranging from 0.45 to 0.55 is shown in Figure 2. Note that, since the MLE $\hat{F}_{n,\beta}$ depends on the ranks of the $T_i - \beta'_0 X_i$, both curves are piecewise constant where jumps are possible if the ordering in $T_i - \beta'_0 X_i$ changes when β changes. Due to the discontinuous nature of the profiled log likelihood and the score function, the estimators are not unique. The result of Theorem 4.1 is valid for any $\hat{\beta}_n$ satisfying (4.2).

4.2. *Efficient estimates involving the MLE $\hat{F}_{n,\beta}$.* Let K be a probability density function with derivative K' satisfying

(K1) The probability density K has support $[-1, 1]$, is twice continuously differentiable and symmetric on \mathbb{R} .

Let $h > 0$ be a smoothing parameter and K_h respectively K'_h be the scaled versions of K and K' respectively, given by

$$K_h(\cdot) = h^{-1}K(h^{-1}(\cdot)) \quad \text{and} \quad K'_h(\cdot) = h^{-2}K'(h^{-1}(\cdot)).$$

The triweight kernel is used in the simulation examples given in the remainder of the paper. Define the density estimate,

$$(4.3) \quad f_{nh,\beta}(t - \beta'_0 x) = \int K_h(t - \beta'_0 x - w) d\hat{F}_{n,\beta}(w).$$

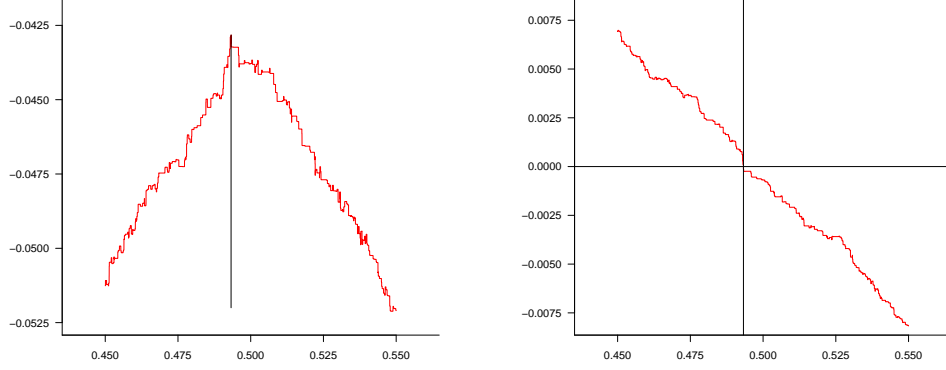


Fig 2: The truncated profile log likelihood $l_n^{(\epsilon)}$ for the MLE $\hat{F}_{n,\beta}$ (left panel) and the score function $\psi_{1,n}^{(\epsilon)}$ (right panel) as a function of β for a sample of size $n = 1000$ and $\epsilon = 0.001$.

We consider

$$(4.4) \quad \psi_{2,nh}^{(\epsilon)}(\beta) \stackrel{\text{def}}{=} \int_{\hat{F}_{n,\beta}(t-\beta'x) \in [\epsilon, 1-\epsilon]} x f_{nh,\beta}(t - \beta'x) \cdot \frac{\hat{F}_{n,\beta}(t - \beta'x) - \delta}{\hat{F}_{n,\beta}(t - \beta'x) \{1 - \hat{F}_{n,\beta}(t - \beta'x)\}} d\mathbb{P}_n(t, x, \delta),$$

and let $\hat{\beta}_n$ be the estimate of β_0 , defined by

$$(4.5) \quad \psi_{2,nh}^{(\epsilon)}(\hat{\beta}_n) = 0.$$

THEOREM 4.2. *Suppose that the conditions of Theorem 4.1 hold and that the function F_β is three times continuously differentiable on the interior of the support S_β . Let $\hat{\beta}_n$ be defined by (4.5). Then, as $n \rightarrow \infty$, and $h \asymp n^{-1/7}$, we have, $\forall \eta > 0$, $\exists n_0 \in \mathbb{N}$ such that $\forall n > n_0$:*

$$\mathbb{P} \left(\exists \beta \in \{b \in \Theta : \|b - \beta_0\| < \eta\} : \psi_{1,n}^{(\epsilon)}(\beta) = 0 \right) > 1 - \eta.$$

Furthermore, we have, $\hat{\beta}_n \xrightarrow{p} \beta_0$, and $\sqrt{n}\{\hat{\beta}_n - \beta_0\}$ converges in distribution to a $N(0, I_\epsilon(\beta_0)^{-1})$ where

$$(4.6) \quad I_\epsilon(\beta_0) = \mathbb{E}_\epsilon \left\{ \frac{f_0(T - \beta_0'X)^2 \{X - \mathbb{E}(X|T - \beta_0'X)\} \{X - \mathbb{E}(X|T - \beta_0'X)\}'}{F_0(T - \beta_0'X) \{1 - F_0(T - \beta_0'X)\}} \right\},$$

assumed to be non-singular.

A picture of the score function $\psi_{2,nh}^{(\epsilon)}(\beta)$ is shown in Figure 3. Note that the range on the vertical axis is considerably larger than the range on the vertical axis of the corresponding Figure 2 for the first method.

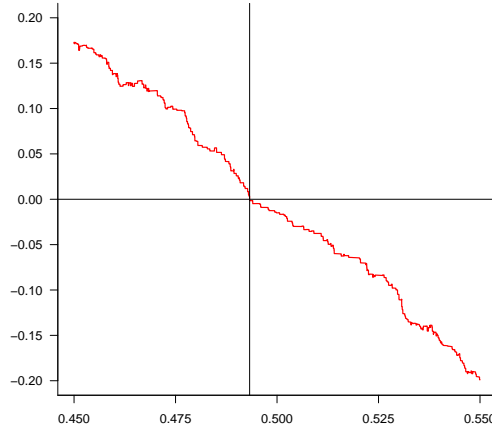


Fig 3: The score function $\psi_{2,nh}^{(\epsilon)}$ as a function of β for a sample of size $n = 1000$ with $\epsilon = 0.001$ and $h = 0.5n^{-1/7}$.

4.3. *Efficient estimates not involving the MLE $\hat{F}_{n,\beta}$.* Define the plug-in estimate

$$(4.7) \quad F_{nh,\beta}(t - \beta'x) = \frac{\int \delta K_h(t - \beta'x - u + \beta'y) d\mathbb{P}_n(u, y, \delta)}{\int K_h(t - \beta'x - u + \beta'y) d\mathbb{G}_n(u, y)},$$

where \mathbb{G}_n is the empirical distribution function of the pairs (T_i, X_i) and where K_h is a scaled version of a probability density function K , satisfying condition (K1); the probability measure of (T, X) will be denoted by G . The plug-in estimates are not necessarily monotone but we show in Theorem 4.4 that $F_{nh,\beta}$ is monotone with probability tending to one as $n \rightarrow \infty$ and $\beta \rightarrow \beta_0$. Another way of writing $F_{nh,\beta}$ is in terms of ordinary sums. Let

$$(4.8) \quad g_{nh,1,\beta}(t - \beta'x) = \frac{1}{n} \sum_{j=1}^n \Delta_j K_h(t - \beta'x - T_j + \beta'X_j),$$

and,

$$(4.9) \quad g_{nh,\beta}(t - \beta'x) = \frac{1}{n} \sum_{j=1}^n K_h(t - \beta'x - T_j + \beta'X_j),$$

then,

$$F_{nh,\beta}(t - \beta'x) = \frac{g_{nh,1,\beta}(t - \beta'x)}{g_{nh,\beta}(t - \beta'x)} = \frac{\sum_{j=1}^n \Delta_j K_h(t - \beta'x - T_j + \beta'X_j)}{\sum_{j=1}^n K_h(t - \beta'x - T_j + \beta'X_j)},$$

in which we recognize the Nadaraya-Watson statistic. One could also omit the diagonal term $j = i$ in the sums above when estimating $F_{nh,\beta}(T_i - \beta'X_i)$ which is often done in the econometric literature (see e.g. [17]). In our computer experiments however, this gave an estimate of the distribution function which had a more irregular behavior than the estimator with the diagonal term included.

If we replace F in (2.3) by $F_{nh,\beta}$, the truncated log likelihood becomes a function of β only. Although the log likelihood has discontinuities if we consider the lower and upper boundaries $F_{nh,\beta}^{-1}(\epsilon)$ and $F_{nh,\beta}^{-1}(1 - \epsilon)$ of the integral also as a function of β , an asymptotic representation of the partial derivatives of the truncated log likelihood is given by the score function,

$$(4.10) \quad \psi_{3,nh}^{(\epsilon)}(\beta) \stackrel{\text{def}}{=} \int_{F_{nh,\beta}(t - \beta'x) \in [\epsilon, 1 - \epsilon]} \frac{\partial_\beta F_{nh,\beta}(t - \beta'x)}{\delta - F_{nh,\beta}(t - \beta'x)} \cdot \frac{d\mathbb{P}_n(t, x, \delta)}{F_{nh,\beta}(t - \beta'x)\{1 - F_{nh,\beta}(t - \beta'x)\}},$$

where the partial derivative of the plug-in estimate $F_{nh,\beta}(t - \beta'x)$, given by (4.7), w.r.t. β has the following form:

$$(4.11) \quad \partial_\beta F_{nh,\beta}(t - \beta'x) = \frac{\int (y - x)\{\delta - F_{nh,\beta}(t - \beta'x)\}K'_h(t - \beta'x - u + \beta'y) d\mathbb{P}_n(u, y, \delta)}{g_{nh,\beta}(t - \beta'x)},$$

where $g_{nh,\beta}(t - \beta'x)$ is defined in (4.9). We define the plug-in estimator $\hat{\beta}_n$ of β_0 by

$$\psi_{3,nh}^{(\epsilon)}(\hat{\beta}_n) = 0.$$

A picture of the truncated log likelihood $l_n^{(\epsilon)}(\beta, F_{nh,\beta})$ and score function $\psi_{3,nh}^{(\epsilon)}(\beta)$ for the plug-in method is shown in Figure 4.

Our main result on the plug-in estimator is given below.

THEOREM 4.3. *If Assumptions (A1)-(A5) hold and*

$$(4.12) \quad -(\beta - \beta_0)' \int_{F_\beta(t - \beta'x) \in [\epsilon, 1 - \epsilon]} \partial_\beta F_\beta(t - \beta'x) \frac{F_0(t - \beta'_0x) - F_\beta(t - \beta'x)}{F_\beta(t - \beta'x)\{1 - F_\beta(t - \beta'x)\}} dG(t, x)$$

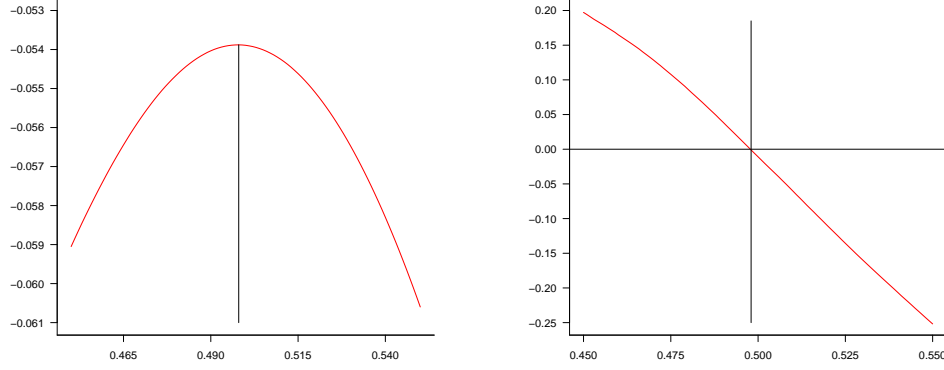


Fig 4: The truncated profile log likelihood $l_n^{(\epsilon)}$ for the plug-in $F_{nh,\beta}$ (left panel) and the score function $\psi_{3,nh}^{(\epsilon)}$ (right panel) as a function of β for a sample of size $n = 1000$ with $\epsilon = 0.001$ and $h = 0.5n^{-1/5}$.

is nonzero for each $\beta \in \Theta$ except for $\beta = \beta_0$. Then, for $\hat{\beta}_n$ being the plug-in estimator introduced above, as $n \rightarrow \infty$, and $h \asymp n^{-1/5}$, we have, $\forall \eta > 0$, $\exists n_0 \in \mathbb{N}$ such that $\forall n > n_0$:

$$\mathbb{P} \left(\exists \beta \in \{b \in \Theta : \|b - \beta_0\| < \eta\} : \psi_{1,n}^{(\epsilon)}(\beta) = 0 \right) > 1 - \eta.$$

Furthermore, we have, $\hat{\beta}_n \xrightarrow{p} \beta_0$ and $\sqrt{n}\{\hat{\beta}_n - \beta_0\}$ is asymptotically normal with mean zero and variance $I_\epsilon(\beta_0)$ where $I_\epsilon(\beta_0)$, defined in (4.6), is assumed to be non-singular.

REMARK 4.2. Note that using an expansion in $\beta - \beta_0$, we can write $\partial_\beta F_\beta(t - \beta'x)$ as,

$$\begin{aligned} & \int (y - x) f_0(t - \beta'_0 x + (\beta - \beta_0)'(y - x)) f_{X|T-\beta'X}(y|T - \beta'X = t - \beta'x) dy \\ & + \int F_0(t - \beta'_0 x + (\beta - \beta_0)'(y - x)) \partial_\beta f_{X|T-\beta'X}(y|T - \beta'X = t - \beta'x) dy \\ & = f_0(t - \beta'x) \mathbb{E} \{X - x|T - \beta'X = t - \beta'x\} + O(\beta - \beta_0) \end{aligned}$$

so that the integral defined in (4.12) can be approximated by,

$$\begin{aligned}
& -(\beta - \beta_0)' \int_{F_\beta(u) \in [\epsilon, 1-\epsilon]} f_0(u) \mathbb{E} \{X - x | T - \beta'X = u\} \\
& \quad \cdot \frac{F_0(u + (\beta - \beta_0)'x) - F_\beta(u)}{F_\beta(u)\{1 - F_\beta(u)\}} f_{X|T-\beta'X}(x|u) dx du \\
& = \int_{F_\beta(u) \in [\epsilon, 1-\epsilon]} \frac{f_0(u) \text{covar}((\beta - \beta_0)'X, F_0(u + (\beta - \beta_0)'X) | T - \beta'X = u)}{F_\beta(u)\{1 - F_\beta(u)\}} du,
\end{aligned}$$

which is positive by the monotonicity of F_0 . (See also the discussion in [23] about this covariance.) A crucial property of the covariance used here, showing that the covariance is nonnegative, goes back to a representation of the covariance in [18], which can easily be proved by an application of Fubini's theorem:

$$EXY - EXEY = \int \{\mathbb{P}(X \geq s, Y \geq t) - \mathbb{P}(X \geq s)\mathbb{P}(Y \geq t)\} ds dt.$$

if XY , X and Y are integrable. Figure 5 shows the integral in (4.12) for our simulation model for $\beta \in [0.45, 0.55]$ and illustrates that this integral is strictly positive except for $\beta = \beta_0$, which is a crucial property for the proof of the consistency of the plug-in estimator given in the supplemental article [12].

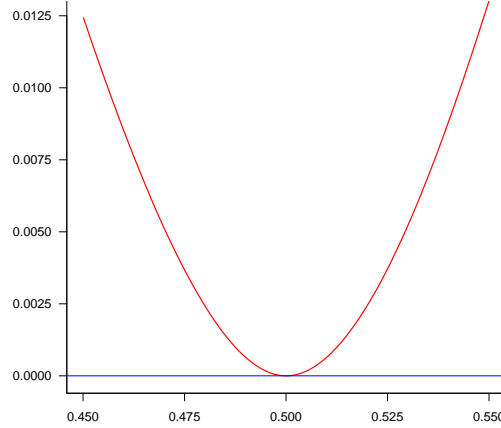


Fig 5: The integral defined in (4.12), as a function of β , with $\epsilon = 0.001$.

Section 4.3.1 contains a road map of the proof of Theorem 4.3, the proof itself is given in the supplemental article [12]. We also have the following results for the plug-in estimate.

THEOREM 4.4. *Let the conditions of Theorem 4.3 be satisfied, then we have on each interval I contained in the support of f_β and for each $\beta \in \Theta$*

$$P \{F_{nh,\beta} \text{ is monotonically increasing on } I\} \xrightarrow{p} 1.$$

The proof of Theorem 4.4 follows from the asymptotic monotonicity of the plug-in estimate in the classical current status model (without regression parameters) and is proved in the same way as Theorem 3.3 of [13].

THEOREM 4.5. *Let the conditions of Theorem 4.3 be satisfied. Then, for $\hat{\beta}_n$ being the plug-in estimator of β_0 ,*

$$\begin{aligned} \sqrt{n}(\hat{\beta}_n - \beta_0) &= \frac{I_\epsilon(\beta_0)^{-1}}{\sqrt{n}} \sum_{i \in J_{F_0}} f_0(T_i - \beta'_0 X_i) \{ \mathbb{E}(X_i | T_i - \beta'_0 X_i) - X_i \} \\ &\quad \cdot \frac{\Delta_i - F_0(T_i - \beta'_0 X_i)}{F_0(T_i - \beta'_0 X_i) \{1 - F_0(T_i - \beta'_0 X_i)\}} + o_p(1). \end{aligned}$$

where $J_H = \{i : \epsilon \leq H(T_i - \beta'_0 X_i) \leq 1 - \epsilon\}$ for some function H .

The representation of Theorem 4.5 plays an important role in determining the variance of smooth functionals, of which the intercept $\alpha = \int u dF_0(u)$ is an example. The proof of Theorem 4.5 is given in the supplemental article [12]. A similar representation holds for the estimators defined in (4.2) and (4.5) (see the proofs of Theorem 4.1 and 4.2 respectively).

REMARK 4.3. The plug-in method also suggests the use of U-statistics. By straightforward calculations, we can write the score function defined in (4.10) as

$$\begin{aligned} \psi_{3,nh}^{(\epsilon)}(\beta) &= \frac{1}{n^2} \sum_{i \in J_{F_{nh,\beta}}} \frac{\frac{\partial}{\partial \beta} F_{nh,\beta}(T_i - \beta' X_i) \{ \Delta_i - F_{nh,\beta}(T_i - \beta' X_i) \}}{F_{nh,\beta}(T_i - \beta' X_i) \{1 - F_{nh,\beta}(T_i - \beta' X_i)\}} \end{aligned}$$

$$\begin{aligned}
&= \frac{1}{n^2} \sum_{i \in J_{F_{nh,\beta}}} \sum_{j \neq i} \frac{\Delta_i \Delta_j (X_j - X_i) K'_h(T_i - \beta' X_i - T_j + \beta' X_j)}{g_{nh,1,\beta}(T_i - \beta' X_i)} \\
&\quad + \frac{1}{n^2} \sum_{i \in J_{F_{nh,\beta}}} \sum_{j \neq i} \frac{(1 - \Delta_i)(1 - \Delta_j)(X_j - X_i) K'_h(T_i - \beta' X_i - T_j + \beta' X_j)}{g_{nh,0,\beta}(T_i - \beta' X_i)} \\
(4.13) \quad &- \frac{1}{n^2} \sum_{i \in J_{F_{nh,\beta}}} \sum_{j \neq i} \frac{(X_j - X_i) K'_h(T_i - \beta' X_i - T_j + \beta' X_j)}{g_{nh,\beta}(T_i - \beta' X_i)}
\end{aligned}$$

where $g_{nh,0,\beta} = g_{nh,\beta} - g_{nh,1,\beta}$, see (4.8) and (4.9). Each of the three terms on the right-hand side of (4.13) can be rewritten in terms of a scaled second order U-statistics. A proof based on U-statistics requires lengthy and tedious calculations which are avoided in the current approach for proving Theorem 4.3. The representation given in Theorem 4.5 also indicates that the U-statistics representation does not give the most natural approach to the proof of asymptotic normality and efficiency of $\hat{\beta}_n$. For these reasons, we do not further examine the results on U-statistics.

REMARK 4.4. We propose the bandwidths $h \asymp n^{-1/7}$ resp. $h \asymp n^{-1/5}$ in Theorem 4.2 resp. Theorem 4.3, which are the usual bandwidths with ordinary second order kernels for the estimates of a density resp. distribution function. Unfortunately, various advices are given in the literature on what smoothing parameters one should use. [21] has fourth order kernels and uses bandwidths between the orders $n^{-1/6}$ and $n^{-1/8}$ for the estimation of F . Note that the use of fourth order kernels needs the associated functions to have four derivatives in order to have the desired bias reduction. [5] advises a bandwidth h such that $n^{-1/5} \ll h \ll n^{-1/8}$, excluding the choice $h \asymp n^{-1/5}$. Both ranges are considerably large and exclude our bandwidth choice $h \asymp n^{-1/5}$. [24] considers a penalized maximum likelihood estimator where the penalty parameter λ_n satisfies $1/\lambda_n = O_p(n^{2/5})$ and $\lambda_n^2 = o_p(n^{-1/2})$. Translated into bandwidth choice (using $h_n \asymp \sqrt{\lambda_n}$), the conditions correspond to: $n^{-1/5} \lesssim h \ll n^{-1/8}$, suggesting that their conditions do allow the choice $h \asymp n^{-1/5}$ for estimating the distribution function.

4.3.1. *Road map of the proof of Theorem 4.3.* The older proofs of a result of this type always used second derivative calculations. As convincingly argued in [33], proofs of this type should only use first derivatives and that is indeed what we do. The limit function F_β of the estimates for F_0 when $\beta \neq \beta_0$ plays a crucial role in our proofs. We first prove the consistency of the plug-in estimate $\hat{\beta}_n$. Next, we use a Donsker property for the functions representing the score function and prove that the integral w.r.t. $d\mathbb{P}_n$ of this score function is

$$o_p\left(n^{-1/2} + \hat{\beta}_n - \beta_0\right),$$

and that the integral w.r.t. dP_0 is asymptotically equivalent to

$$-(\hat{\beta}_n - \beta_0)I_\epsilon(\beta_0),$$

where $I_\epsilon(\beta_0)$ is the generalized Fisher information, given by (4.6). Combining these results gives Theorem 4.3. Very essential in this proof are L_2 -bounds on the distance between the functions $F_{nh,\beta}$ to its limit F_β for fixed β and on the L_2 -distance between the first partial derivatives $\partial_\beta F_{nh,\beta}$ and $\partial_\beta F_\beta$. If the bandwidth $h \asymp n^{-1/5}$, the first L_2 -distance is of order $n^{-2/5}$ and the second distance is of order $n^{-1/5}$, allowing us to use the Cauchy-Schwarz inequality on these components. Here we use a result in [11] on L_2 bounds of derivatives of kernel density estimates.

In Section 5 we discuss the estimation of an efficient estimate of the intercept term in regression model (2.1) using the plug-in estimates $\hat{\beta}_n$ and $F_{nh,\hat{\beta}_n}$.

4.4. Truncation. We introduced a truncation device in order to avoid unbounded score functions and numerical difficulties. If one starts with the *efficient* score equation or an estimate thereof, the solution sometimes suggested in the literature, is to add a constant c_n , tending to zero as $n \rightarrow \infty$, to the factor $F(t - \beta'x)\{1 - F(t - \beta'x)\}$ which inevitably will appear in the denominator. This is done in, e.g. [23]; similar ideas involving a sequence (c_n) are used in [21] and [5].

In contrast with the usual approaches to truncation, which imply the selection of a suitable sequence c_n , we do not consider a vanishing truncation sequence but work with a subsample of the data depending on the ϵ and $(1 - \epsilon)$ quantiles of the distribution function estimate for small but fixed $\epsilon \in (0, 1/2)$. This simple device in (2.3) moreover implies keeping the characterizing properties of the MLE (see Proposition 1.1 on p. 39 of [15]) which are lost when a vanishing sequence is considered. It is perhaps somewhat remarkable that we can, instead of letting $\epsilon \downarrow 0$, fix $\epsilon > 0$ and still have consistency of our estimators; on the other hand, the estimate proposed by [24] is also identified via a subset of the support of the distribution F_0 .

Although the truncation area depends on β , we show in the supplemental article [12] (see the proof of Theorem 4.1) that the population version of the score function, given by

$$(4.14) \quad \psi_\epsilon(\beta) = \int_{F_\beta(t - \beta'x) \in [\epsilon, 1 - \epsilon]} \phi(t, x, \delta) \{F_\beta(t - \beta'x) - \delta\} dP_0(t, x, \delta),$$

has a derivative at $\beta = \beta_0$ that only involves the derivative of the integrand in (4.14), but does not involve terms arising from the truncation limits appearing in the integral. Using the truncation in the argmax maximum log likelihood approach would not lead to a derivative of the population version of the log likelihood which ignores the boundaries and therefore this truncation is less suited for argmax estimators.

A drawback of our fixed truncation parameter approach is that we get a truncated Fisher information. The resulting estimates are therefore not efficient in the classical sense of efficiency but the difference between the efficient variance and almost (determined by the size of ϵ) efficient variance is rather small in our simulation models. We also tried to program the fully efficient estimators proposed by [23] and compared its performance to the performance of our almost-efficient estimators. The comparison showed that our estimates perform better in finite samples. Moreover, the estimates by [23] involve several kernel density estimates, resulting in a very large computation time compared to our simple estimates (involving 5 double summations over the data points).

Moreover, the usual conditions in the theory of estimation of F_0 under current status and, more generally, interval censored data are that F_0 corresponds to a distribution with compact support. Otherwise, certain variances easily get infinite, and similarly, the Fisher information in our model can easily become infinite. Truncating by keeping the quantiles between ϵ and $1 - \epsilon$ avoids difficulties in this case and allows us to apply the theory which presently has been developed for the current status model.

Note that the score function defined in (4.1) does not contain a factor $F(t - \beta'x)$ or $1 - F(t - \beta'x)$ in the denominator. For simplicity of the proofs, we still impose the truncation area, since the classical results for the current status model are derived under the assumption that the density f_0 is bounded away from zero. We conjecture however that the result of Theorem 4.1 remains valid when taking $\epsilon = 0$.

5. Estimation of the intercept. We want to estimate the intercept

$$(5.1) \quad \alpha = \int u dF_0(u).$$

We can take the plug-in estimate $\hat{\beta}_n$ of β_0 , by using a bandwidth of order $n^{-1/5}$ and the score procedure, as before. However, in estimating α , as defined by (5.1), we have to estimate F_0 with a smaller bandwidth h , satisfying $h \ll n^{-1/4}$ to avoid bias, for example $h \asymp n^{-1/3}$. The matter is discussed in [5], p. 1253.

We have the following result of which the proof can be found in the supplemental article [12].

THEOREM 5.1. *Let the conditions of Theorem 4.3 be satisfied, and let $\hat{\beta}_n$ be the k -dimensional estimate of β_0 as obtained by the score procedure, described in Theorem 4.3, using a bandwidth of order $n^{-1/5}$. Let $F_{nh, \hat{\beta}_n}$ be a plug-in estimate of F_0 , using $\hat{\beta}_n$ as the estimate of β_0 , but using a bandwidth h of order $n^{-1/3}$ instead of $n^{-1/5}$. Finally, let $\hat{\alpha}_n$ be the estimate of α , defined by*

$$\int u dF_{nh, \hat{\beta}_n}(u).$$

Then $\sqrt{n}(\hat{\alpha}_n - \alpha)$ is asymptotically normal, with expectation zero and variance

$$(5.2) \quad \sigma^2 \stackrel{\text{def}}{=} a(\beta_0)' I_\epsilon(\beta_0)^{-1} a(\beta_0) + \int \frac{F_0(v)\{1 - F_0(v)\}}{f_{T-\beta_0'X}(v)} dv,$$

where $a(\beta_0)$ is the k -dimensional vector, defined by

$$a(\beta_0) = \int E_{\beta_0}\{X|T - \beta_0'X = u\} f_0(u) du,$$

and $I_\epsilon(\beta_0)$ is defined in (4.6).

REMARK 5.1. We choose the bandwidth of order $n^{-1/3}$ for specificity, but other choices are also possible. We can in fact choose $n^{-1/2} \ll h \ll n^{-1/4}$. The bandwidth of order $n^{-1/3}$ corresponds to the automatic bandwidth choice of the MLE of F_0 , also using the estimate $\hat{\beta}_n$ of β_0 .

REMARK 5.2. Note that the variance corresponds to the information lower bound for smooth functionals in the binary choice model, given in [5]. The second part of the expression for the variance on the right-hand side of (5.2) is familiar from current status theory, see e.g. (10.7), p. 287 of [14].

Instead of considering the plug-in estimate, we could also consider the estimates described in Theorem 4.1 and Theorem 4.2. After having determined an estimate $\hat{\beta}_n$ in this way, we next estimate α by

$$(5.3) \quad \hat{\alpha}_n = \int x d\hat{F}_{n,\hat{\beta}_n}(x),$$

where $\hat{F}_{n,\hat{\beta}_n}$ is the MLE corresponding to the estimate $\hat{\beta}_n$. The theoretical justification of this approach can be proved using the asymptotic theory of smooth functionals given in [14], p.286. Using the MLE $\hat{F}_{n,\hat{\beta}_n}$ instead of the plug-in $\hat{F}_{nh,\hat{\beta}_n}$ as an estimate of the distribution function F_0 , avoids the selection of a bandwidth parameter for the intercept estimate. We discuss in the next section how the bandwidth can be selected by the practitioner in a real data sample.

6. Computation and simulation. The computation of our estimates is relatively straightforward in all cases. For the score-estimates defined in Sections 4.1 and 4.2, we first compute the MLE for fixed β by the so-called “pool adjacent violators” algorithm for computing the convex minorant of the “cusum diagram” defined in (3.1). If the MLE has been computed for fixed β , we can compute the density estimate f_{nh} . The estimate of β_0 is then determined by a root-finding algorithm

such as Brent's method. Computation is very fast. For the plug-in estimate, we simply compute the estimate $F_{nh,\beta}$ as a ratio of two kernel estimators for fixed β and then compute the derivative w.r.t. β . Next we use again a root-finding algorithm to determine the zero of the corresponding score function.

Some results from the simulations of our model are available in Table 1, which contains the mean value of the estimate, averaged over $N = 10,000$ iterations, and n times the variance of the estimate of $\beta_0 = 0.5$ (resp. $\alpha_0 = 0.5$) for the different methods described above, as well as for the classical MLE of β_0 , for different sample sizes n and a truncation parameters $\epsilon = 0.001$. We took the bandwidth $h = 0.5n^{-1/7}$ for the efficient score-estimate of Section 4.2. The bandwidth $h = 0.5n^{-1/5}$ for the plug-in estimate of Section 4.3 was chosen based on an investigation of the mean squared error (MSE) for different choices of c in $h = cn^{-1/5}$. Details on how to choose the bandwidth in practice are given in Section 6.1. The true asymptotic values for the variance of $\sqrt{n}(\hat{\beta}_n - \beta_0)$ in our simulation model, obtained via the inverse of the Fisher information $I_\epsilon(\beta_0)$, are 0.151707 without truncation and 0.158699 for $\epsilon = 0.001$ and 0.17596 for $\epsilon = 0.01$. We advise to use a truncation parameter ϵ of 0.001 or smaller in practice. The variance defined in Theorem 4.1 for $\epsilon = 0.001$ is 0.193612. The lower bounds for the variance of the intercept are 0.257898 for the simple score method and 0.222984 for the efficient methods. Our results show slow convergence to these bounds.

Table 1 shows that the efficient score and the efficient plug-in methods perform reasonably well. A drawback of the plug-in method however is the long computing time for large sample sizes, whereas the computation for the MLE is fast even for the larger samples. Note moreover that the plug-in estimate is only asymptotically monotone whereas the MLE is monotone by definition. All our proposed estimates perform better than the classical MLE, the log likelihood for the MLE has moreover a rough behavior, with a larger chance that optimization algorithms might calculate a local maximizer instead of the global maximizer.

The performance of the score estimates is worse than the performance of the plug-in estimates for small sample sizes but increases considerably when the sample size increases. Although the asymptotic variance of the first score-estimator of Section 4.1 is larger than the (almost, determined by the truncation parameter ϵ) efficient variance, the results obtained with this method are noteworthy seen the fact that no smoothing is involved in this simple estimation technique.

Table 1 does not provide strong evidence of the \sqrt{n} -consistency of the classical MLE, but we conjecture that the MLE is indeed \sqrt{n} -consistent but not efficient. Considering the drawbacks of the classical MLE, we advise the use of the plug-in estimate for small sample sizes and the use of the score estimates, based on the MLE, for larger sample sizes, for estimating the parameter β_0 . We finally suggest to estimate the parameter α_0 via the MLE corresponding to this β_0 -estimate, avoiding

in this way the bias problem for the kernel estimates of α_0 .

TABLE 1

The mean value of the estimate and n times the variance of the estimates of β_0 and α_0 for different methods, $h_\beta = 0.5n^{-1/7}$ (for the efficient score method), $h_\beta = 0.5n^{-1/5}$ and $h_\alpha = 0.75n^{-1/3}$ (for the plug-in method), $\epsilon = 0.001$ and $N = 10000$.

		Score-1		Score-2		Plugin		MLE	
	n	mean	$n \times \text{var}$	mean	$n \times \text{var}$	mean	$n \times \text{var}$	mean	$n \times \text{var}$
β	100	0.500212	0.364558	0.502247	0.410449	0.499562	0.245172	0.489690	0.307961
	500	0.499845	0.221484	0.499825	0.230178	0.498857	0.191857	0.499315	0.228335
	1000	0.499982	0.211608	0.500353	0.208102	0.499502	0.192223	0.499937	0.228420
	5000	0.499901	0.195294	0.499964	0.184807	0.500314	0.181421	0.499933	0.239898
	10000	0.499988	0.191115	0.499985	0.172758	0.500120	0.172043	0.499994	0.227222
	20000	0.500038	0.187616	0.500023	0.169762	0.500096	0.174197	0.499952	0.238400
α	100	0.511937	0.468415	0.509679	0.515638	0.495709	0.332949	0.523103	0.425614
	500	0.502258	0.293585	0.502506	0.287576	0.498932	0.254040	0.502514	0.304540
	1000	0.500839	0.284958	0.500616	0.262684	0.498385	0.270085	0.500937	0.300201
	5000	0.500345	0.262566	0.500316	0.244892	0.501597	0.241294	0.500270	0.303754
	10000	0.500127	0.256983	0.500134	0.232973	0.501680	0.245993	0.500076	0.289905
	20000	0.500020	0.250720	0.500042	0.230901	0.501660	0.244042	0.500101	0.302824

6.1. *Bandwidth selection.* In this section we discuss the bandwidth selection for the plug-in estimate. A similar idea can be used for the selection of the bandwidth used for the second estimate defined in Section 4.2. We define the optimal constant c_{opt} in $h = cn^{-1/5}$ as the minimizer of MSE ,

$$c_{opt} = \arg \min_c MSE(c) = \arg \min_c E_{\beta_0} (\hat{\beta}_{n,h_c} - \beta_0)^2,$$

where $\hat{\beta}_{n,h_c}$ is the estimate obtained when the constant c is chosen in the estimation method. A picture of the Monte Carlo estimate of MSE as a function of c is shown for the plug-in method in Figure 6, where we estimated $MSE(c)$ on a grid $c = 0.01, 0.05, 0.10, \dots, 0.95$, for a sample size $n = 1000$ and truncation parameter $\epsilon = 0.001$ by a Monte Carlo experiment with $N = 1000$ simulation runs,

$$(6.1) \quad \widehat{MSE}(c) = N^{-1} \sum_{j=1}^N (\hat{\beta}_{n,h_c}^j - \beta_0)^2,$$

where $\hat{\beta}_{n,h_c}^j$ is the estimate of β_0 in the j -th simulation run, $j = 1, \dots, N$.

Since F_0 and β_0 are unknown in practice, we cannot compute the actual MSE . We use the bootstrap method proposed by [29] to obtain an estimate of MSE . Our proposed estimate $F_{nh,\beta}$ of the distribution function F_0 satisfy the conditions of Theorem 3 in [29] and the consistency of the bootstrap is guaranteed. Note that it follows from [22] and [28] that naive bootstrapping, by resampling with replacement

(T_i, X_i, Δ_i) , or by generating bootstrap samples from the MLE, is inconsistent for reproducing the distribution of the MLE.

The method works as follows. We let $h_0 = c_0 n^{-1/5}$ be an initial choice of the bandwidth and calculate the plug-in estimates $\hat{\beta}_{n,h_0}$ and $F_{n,h_{c_0}}$ based on the original sample $(T_i, X_i, \Delta_i), i = 1, \dots, n$. We generate a bootstrap sample $(X_i, T_i, \Delta_i^*), i = 1, \dots, n$ where the (T_i, X_i) correspond to the (T_i, X_i) in the original sample and where the indicator Δ_i^* is generated from a Bernoulli distribution with probability $F_{n,h_0}(T_i - \hat{\beta}_{n,h_0} X_i)$, and next estimate $\hat{\beta}_{n,h_c}^*$ from this bootstrap sample. We repeat this B times and estimate $MSE(c)$ by,

$$(6.2) \quad \widehat{MSE}_B(c) = B^{-1} \sum_{b=1}^B (\hat{\beta}_{n,h_c}^{*b} - \hat{\beta}_{n,h_{c_0}})^2,$$

where $\hat{\beta}_{n,h_c}^{*b}$ is the bootstrap estimate in the b -th bootstrap run. The optimal bandwidth $\hat{h}_{opt} = \hat{c}_{opt} n^{-1/5}$ where \hat{c}_{opt} is defined as the minimizer of $\widehat{MSE}_B(c)$.

To analyze the behavior of the bootstrap method, we compared the Monte Carlo estimate of MSE , defined in (6.1), (based on $N = 1000$ samples of size $n = 1000$) to the bootstrap MSE defined in (6.2) (based on a single sample of size $n = 1000$) in Figure 6. The figure shows that the Monte Carlo MSE and the bootstrap MSE are in line, which illustrates the consistency of the method. The choice of the initial bandwidth does effect the size of the estimated MSE but not the behavior of the estimate and we conclude that this bootstrap algorithm can be used to select an optimal bandwidth parameter in the described method above.

7. Discussion. In this paper we propose a simple \sqrt{n} -consistent estimate for the finite dimensional regression parameter in the semi-parametric current status linear regression model with unspecified error distribution. The estimate has an asymptotically normal limiting distribution but does not attain the efficiency bounds. We also consider two different methods to obtain \sqrt{n} -consistent and asymptotic normal estimates with an asymptotic variance that is arbitrarily close to the efficient variance. The first approach uses the MLE for the distribution function F for fixed β , the second approach does not depend on the behavior of this MLE but uses a kernel estimate for the distribution function. All proposed estimates are defined as the root of a score function as a function of β .

We introduced a truncation device to avoid theoretical and numerical difficulties caused by unbounded score functions. The truncation is carried out by considering a subsample of the data depending on the ϵ and $1 - \epsilon$ quantiles of the distribution function estimate. Although our estimates do not attain the efficiency bound, the proposed method allows for easy computation of the estimates without the need for selecting a truncation sequence converging to zero. Achieving efficiency at the cost

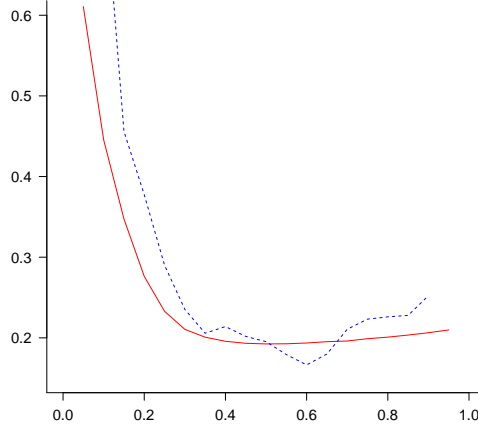


Fig 6: Estimated $MSE(c)$ plot of $\hat{\beta}_n$ obtained from $N = 1000$ Monte Carlo simulations (red, solid) versus the bootstrap MSE for $c_0 = 0.25$ (blue, dashed) with $B = 10000$, $n = 1000$ and $\epsilon = 0.001$.

of additional computational complexities associated with smoothing procedures and truncation sequence selection results in only a small asymptotic efficiency gain and does not seem to improve the performance of our simple methods.

The estimates based on the efficient score function depending on the MLE for F_0 for fixed β have a slightly better performance than the estimates based on the smooth score function depending on the plug-in estimates for F_0 when the sample size is large. For small samples none of the MLE-based estimates comes out as uniformly best.

8. Appendix. In this section, we include the derivation of the efficient information bound for the current status linear regression model. The proofs of the results given in Section 4.1, Section 4.2 and Section 4.3 are deferred to the supplemental article [12].

8.1. *Efficient information in the current status linear regression model.* The density of one observation in the current status linear regression model is,

$$p_{\beta,F}(t, x, \delta) = F(t - \beta'x)^\delta \{1 - F(t - \beta'x)\}^{1-\delta} f_{T,X}(t, x).$$

We assume that the distribution of (T, X) does not depend on (β, F) which implies that the relevant part of the log likelihood is given by:

$$l_n(\beta, F) = \sum_{i=1}^n [\Delta_i \log F(T_i - \beta' X_i) + (1 - \Delta_i) \log \{1 - F(T_i - \beta' X_i)\}]$$

If the distribution F is known (parametric case), the information for β is given by,

$$I_P(\beta) = E \left(\left(\frac{\partial}{\partial \beta} \log p_{\beta, F}(T, X, \Delta) \right)' \left(\frac{\partial}{\partial \beta} \log p_{\beta, F}(T, X, \Delta) \right) \right).$$

Straightforward calculations yield,

$$I_P(\beta)_{ij} = \int \frac{\mathbb{E}(X_i X_j | T - \beta' X = u)}{F(u) \{1 - F(u)\}} f(u)^2 f_{T - \beta' X}(u) du,$$

where $f = F'$ and where $f_{T - \beta' X}$ is the density of $T - \beta' X$. When F is unknown, we need to calculate the efficient score function. Let F and P_0 be the probability measures of ε and (T, X, Δ) respectively and let $L_2^0(Q)$ be the Hilbert space of square integrable functions a with respect to the measure dQ satisfying $\int a dQ = 0$. The score operator $l_F : L_2^0(F) \mapsto L_2^0(P_0)$ is defined by,

$$\begin{aligned} [l_F a](t, x, \delta) &= E(a(\varepsilon) | (T, X, \Delta) = (t, x, \delta)) \\ &= \frac{\delta \int_{-\infty}^{t - \beta' x} a(s) dF(s)}{F(t - \beta' x)} - \frac{(1 - \delta) \int_{-\infty}^{t - \beta' x} a(s) dF(s)}{1 - F(t - \beta' x)} \end{aligned}$$

with adjoint,

$$[l_F^* b](e) = E(b(T, X, \Delta) | \varepsilon = e).$$

The information for β in the semi-parametric model is defined by,

$$I(\beta) = E \left(\tilde{\ell}_{\beta, F}(T, X, \Delta)' \tilde{\ell}_{\beta, F}(T, X, \Delta) \right)$$

where $\tilde{\ell}_{\beta, F}(t, x, \delta)$ is the efficient score function defined by,

$$\tilde{\ell}_{\beta, F}(t, x, \delta) = \ell_{\beta}(t, x, \delta) - [l_F a_*](t, x, \delta),$$

where

$$\ell_{\beta}(t, x, \delta) = \frac{\partial}{\partial \beta} \log p_{\beta, F}(t, x, \delta) = \frac{-\delta x f(t - \beta' x)}{F(t - \beta' x)} + \frac{(1 - \delta) x f(t - \beta' x)}{1 - F(t - \beta' x)},$$

and $\ell_F a_*$ satisfies,

$$(8.1) \quad \ell_F^* \ell_F a_* = \ell_F^* \ell_\beta.$$

The efficient score $\tilde{\ell}_{\beta,F}$ can be interpreted as the residual of ℓ_β projected in the space spanned by $\ell_F a$ for $a \in L_2^0(F)$. Note that, as a consequence of (8.1), the efficient information is

$$I(\beta) = E \left(\tilde{\ell}_{\beta,F}(T, X, \Delta)' \ell_\beta(T, X, \Delta) \right).$$

To find a_* , we have to solve (8.1),

$$\begin{aligned} \ell_F^* \ell_F a_*(e) &= \int_e^\infty \frac{\phi(u)}{F(u)} f_{T-\beta X}(u) du - \int_{-\infty}^e \frac{\phi(u)}{1-F(u)} f_{T-\beta X}(u) du \\ &= - \int_e^{+\infty} \frac{\mathbb{E}(X|T-\beta X=u) f(u)}{1-F(u)} f_{T-\beta X}(u) du \\ &\quad + \int_{-\infty}^e \frac{\mathbb{E}(X|T-\beta X=u) f(u)}{1-F(u)} f_{T-\beta X}(u) du \\ (8.2) \quad &= \ell_F^* \ell_\beta(e), \end{aligned}$$

where $\phi(t) = \int_{-\infty}^t a(s) dF(s)$. Equation (8.2) is satisfied with

$$\phi(u) = -\mathbb{E}(X|T-\beta X=u) f(u).$$

Any a_* that satisfies the above equation satisfies (8.1) and we get,

$$\begin{aligned} \tilde{\ell}_{\beta,F}(t, x, \delta) &= \\ &\left\{ E(X|T-\beta'X=t-\beta'x) - x \right\} f(t-\beta'x) \left\{ \frac{\delta}{F(t-\beta'x)} - \frac{1-\delta}{1-F(t-\beta'x)} \right\}, \end{aligned}$$

and,

$$(8.3) \quad I(\beta)_{ij} = \int \frac{\text{covar}(X_i, X_j|T-\beta'X=u)}{F(u)\{1-F(u)\}} f(u)^2 f_{T-\beta'X}(u) du.$$

Note that $I_P(\beta)^{-1} - I(\beta)^{-1}$ equals the minimal increase of the variance of an estimator for β based on an unknown F (semi-parametric case) compared to the situation where F is known (parametric). In our simulation example $I_P(\beta) = 26.3667$ and $I(\beta) = 6.5917$.

9. Supplementary material. Here we give the proofs of the results stated in Sections 4.1, 4.2 and 4.3 of the manuscript. Entropy results are used in our proofs. Before we prove the results we first give some definitions and an equicontinuity lemma needed in the proofs.

Consider a class of functions \mathcal{F} in $L_2(Q)$ where $L_2(Q)$ is the L_2 seminorm defined by a probability measure Q on \mathcal{R} , i.e. for $g \in \mathcal{F}$,

$$\|g\|_{2,Q} = \left\{ \int g^2 dQ \right\}^{1/2}.$$

For any $\zeta > 0$ let $N_2(\zeta, \mathcal{F}, Q)$ be the minimal number N for which there exists functions g_1, \dots, g_N such that for each $g \in \mathcal{F}$ there is a $j \in \{1, \dots, N\}$ such that $\|g - g_j\|_{2,Q} \leq \zeta$. $N_2(\zeta, \mathcal{F}, Q)$ is called the ζ -covering number and $H_2(\zeta, \mathcal{F}, Q) = \log N_2(\zeta, \mathcal{F}, Q)$ is called the ζ -entropy of \mathcal{F} . We have the following result from [25].

LEMMA S9.1 (Equicontinuity Lemma, p. 150 in [25]). *Let \mathcal{F} be a permissible class of functions with envelope F in $L_2(P) = \{f : \int f^2 dP < \infty\}$. Suppose that for each $\eta > 0$ and $\varepsilon > 0$ there exists a $\gamma > 0$ such that*

$$\limsup_n P \left(\int_0^\gamma \{\log N_2(u, \mathcal{F}, \mathbb{P}_n)/u\}^{1/2} du > \eta \right) < \varepsilon,$$

where \mathbb{P}_n is the empirical probability measure of n observations from P . Then, there exists $\delta > 0$ for which

$$\limsup_{n \rightarrow \infty} P \left(\sup_{[\delta]} |\sqrt{n}| \left| \int (f - g) d(\mathbb{P}_n - P) \right| > \eta \right) < \eta,$$

where,

$$[\delta] = \{(f, g) : f, g \in \mathcal{F} \text{ and } \|f - g\|_{2,P} \leq \delta\}.$$

S10. Asymptotic behavior of the simple estimate based on the MLE $\hat{F}_{n,\beta}$, avoiding any smoothing. This section contains the proof of Theorem 4.1 stated in Section 4.1 of the manuscript. The proof is decomposed into three parts: (a) proof of existence of a root of $\psi_{1,n}$, (b) proof of consistency of $\hat{\beta}_n$ and (c) proof of asymptotic normality of $\sqrt{n}(\hat{\beta}_n - \beta_0)$.

PROOF OF THEOREM 4.1, PART 1 (EXISTENCE OF A ROOT). Consider the score function

$$\psi_{1,n}^{(\epsilon)}(\beta) = \int_{\hat{F}_{n,\beta}(t - \beta'x) \in [\epsilon, 1 - \epsilon]} x \{ \hat{F}_{n,\beta}(t - \beta'x) - \delta \} d\mathbb{P}_n(t, x, \delta),$$

where $\hat{F}_{n,\beta}$ is the nonparametric maximum likelihood estimator (MLE) of the error distribution. Let $\psi_{1,\epsilon}$ be the population version of the statistic $\psi_{1,n}^{(\epsilon)}$ defined by,

$$(S10.1) \quad \psi_{1,\epsilon}(\beta) = \int_{F_\beta(t-\beta'x) \in [\epsilon, 1-\epsilon]} x \{F_\beta(t-\beta'x) - \delta\} dP_0(t, x, \delta).$$

We have

$$\psi_{1,\epsilon}(\beta_0) = 0.$$

We next show that

$$(S10.2) \quad \psi'_{1,\epsilon}(\beta_0) = \mathbb{E}_\epsilon [X \{ \mathbb{E}(X|T - \beta'_0 X) - X \}' f_0(T - \beta'_0 X)]$$

and

$$(S10.3) \quad \psi_{1,n}^{(\epsilon)}(\beta) = \psi'_{1,\epsilon}(\beta_0)(\beta - \beta_0) + R_n,$$

where $R_n = o_p(1) + o(\beta - \beta_0)$. Note that $\psi'_{1,\epsilon}(\beta_0)$ is a weighted expectation of conditional covariance and is by definition non-singular. This implies that,

$$\psi_{1,n}^{(\epsilon)}(\beta) = 0 \iff \beta = \beta_0 - \psi'_{1,\epsilon}(\beta_0)^{-1} R_n,$$

where $R_n = o_p(1) + o(\beta - \beta_0)$. Note that the $o_p(1)$ term depends on the sample, and tends to zero in probability as $n \rightarrow \infty$, whereas the $o(\beta - \beta_0)$ term is non-random and only depends on the parameters of the underlying model.

As a consequence, the function $\beta \mapsto \psi_{1,n}^{(\epsilon)}(\beta)$ has a zero in each small open neighborhood of β_0 with probability tending to 1 i.e., $\forall \eta > 0, \exists n_0 \in \mathbb{N}$ such that $\forall n > n_0 :$

$$\mathbb{P} \left(\exists \beta \in \{b \in \Theta : \|\beta_0 - b\| < \eta\} : \psi_{1,n}^{(\epsilon)}(\beta) = 0 \right) > 1 - \eta.$$

We start by computing the derivative of the population version of the score function at $\beta = \beta_0$. We have,

$$\begin{aligned} \psi'_{1,\epsilon}(\beta) &= \frac{\partial}{\partial \beta} \int_{F_\beta^{-1}(\epsilon) \leq t - \beta'x \leq F_\beta^{-1}(1-\epsilon)} x \{F_\beta(t - \beta'x) - \delta\} dP_0(t, x, \delta) \\ &= \frac{\partial}{\partial \beta} \int_{F_\beta^{-1}(\epsilon) \leq t - \beta'x \leq F_\beta^{-1}(1-\epsilon)} x \{F_\beta(t - \beta'x) - F_0(t - \beta'_0 x)\} dG(t, x) \\ &= \frac{\partial}{\partial \beta} \int_{u=F_\beta^{-1}(\epsilon)}^{F_\beta^{-1}(1-\epsilon)} \int x \{F_\beta(u) - F_0(u + (\beta - \beta_0)'x)\} f_{X|T-\beta'X}(x|u) f_{T-\beta'X}(u) dx du \end{aligned}$$

$$\begin{aligned}
&= \int_{u=F_\beta^{-1}(\epsilon)}^{F_\beta^{-1}(1-\epsilon)} \int \frac{\partial}{\partial \beta} \left\{ x \left\{ F_\beta(u) - F_0(u + (\beta - \beta_0)'x) \right\} f_{X|T-\beta'X}(x|u) f_{T-\beta'X}(u) \right\} dx du \\
&\quad + \left\{ \frac{\partial}{\partial \beta} F_\beta^{-1}(1-\epsilon) \right\} \int x \left\{ 1 - \epsilon - F_0(F_\beta^{-1}(1-\epsilon) + (\beta - \beta_0)'x) \right\} \\
&\quad \quad \cdot f_{X|T-\beta'X}(x|F_\beta^{-1}(1-\epsilon)) f_{T-\beta'X}(F_\beta^{-1}(1-\epsilon)) dx \\
&\quad - \left\{ \frac{\partial}{\partial \beta} F_\beta^{-1}(\epsilon) \right\} \int x \left\{ \epsilon - F_0(F_\beta^{-1}(\epsilon) + (\beta - \beta_0)'x) \right\} \\
&\quad \quad \cdot f_{X|T-\beta'X}(x|F_\beta^{-1}(\epsilon)) f_{T-\beta'X}(F_\beta^{-1}(\epsilon)) dx
\end{aligned}$$

Note that if $\beta = \beta_0$, we get:

$$\begin{aligned}
\psi'_{1,\epsilon}(\beta_0) &= \\
&\int_{F_0^{-1}(\epsilon)}^{F_0^{-1}(1-\epsilon)} \int \frac{\partial}{\partial \beta} \left\{ x \left\{ F_\beta(u) - F_0(u + (\beta - \beta_0)'x) \right\} f_{X|T-\beta'X}(x|u) f_{T-\beta'X}(u) \right\} \Big|_{\beta=\beta_0} du dx
\end{aligned}$$

since the last two terms vanish because the integrands become zero in that case. Note that,

$$\begin{aligned}
\frac{\partial}{\partial \beta} F_\beta(u) &= \int y f_0(u + (\beta - \beta_0)'y) f_{X|T-\beta'X}(y|u) dy \\
&\quad + \int F_0(u + (\beta - \beta_0)'y) \frac{\partial}{\partial \beta} f_{X|T-\beta'X}(y|u) dy,
\end{aligned}$$

implying that, at $\beta = \beta_0$,

$$\frac{\partial}{\partial \beta} F_\beta(u) \Big|_{\beta=\beta_0} = f_0(u) \mathbb{E}(X|T - \beta'_0 X = u).$$

Since

$$\begin{aligned}
&\int_{u=F_0^{-1}(\epsilon)}^{F_0^{-1}(1-\epsilon)} \int x \left\{ \mathbb{E}(X|T - \beta'_0 X = u) - x \right\}' f_{X|T-\beta'_0 X}(x|u) f_0(u) f_{T-\beta'_0 X}(u) dx du \\
&= \mathbb{E}_\epsilon \left[X \left\{ \mathbb{E}(X|T - \beta'_0 X) - X \right\}' f_0(T - \beta'_0 X) \right],
\end{aligned}$$

(S10.2) now follows.

The result of (S10.3) follows by showing that for each $\beta \in \Theta$,

$$\text{(S10.4)} \quad \psi_{1,n}^{(\epsilon)} = \psi_{1,\epsilon}(\beta) + o_p(1),$$

since this implies, using a Taylor expansion of $\psi_{1,\epsilon}$ around $\beta = \beta_0$, that

$$\psi_{1,n}^{(\epsilon)}(\beta) = \psi_{1,\epsilon}(\beta) + o_p(1) = \psi'_{1,\epsilon}(\beta_0)(\beta - \beta_0) + o(\beta - \beta_0) + o_p(1).$$

We have:

(S10.5)

$$\begin{aligned}
\psi_{1,n}^{(\epsilon)}(\beta) &= \int_{\hat{F}_{n,\beta}(t-\beta'x) \in [\epsilon, 1-\epsilon]} x \{F_\beta(t - \beta'x) - \delta\} d\mathbb{P}_n(t, x, \delta) \\
&\quad + \int_{\hat{F}_{n,\beta}(t-\beta'x) \in [\epsilon, 1-\epsilon]} x \{\hat{F}_{n,\beta}(t - \beta'x) - F_\beta(t - \beta'x)\} d\mathbb{P}_n(t, x, \delta) \\
&= \int_{\hat{F}_{n,\beta}(t-\beta'x) \in [\epsilon, 1-\epsilon]} x \{F_\beta(t - \beta'x) - \delta\} d\mathbb{P}_n(t, x, \delta) \\
&\quad + \int_{\hat{F}_{n,\beta}(t-\beta'x) \in [\epsilon, 1-\epsilon]} x \{\hat{F}_{n,\beta}(t - \beta'x) - F_\beta(t - \beta'x)\} d(\mathbb{P}_n - P_0)(t, x, \delta) \\
&\quad + \int_{\hat{F}_{n,\beta}(t-\beta'x) \in [\epsilon, 1-\epsilon]} x \{\hat{F}_{n,\beta}(t - \beta'x) - F_\beta(t - \beta'x)\} dP_0(t, x, \delta).
\end{aligned}$$

Let \mathcal{F} be the set of piecewise constant distribution functions with finitely many jumps (like the MLE $\hat{F}_{n,\hat{\beta}_n}$), and let, for $\beta \in \Theta$, \mathcal{K}_β be the set of functions

$$(S10.6) \quad \mathcal{K}_\beta = \{(t, x, \delta) \mapsto x \{F_\beta(t - \beta'x) - \delta\} 1_{[\epsilon, 1-\epsilon]}(F(t - \beta'x)) : F \in \mathcal{F}\}.$$

We add the function

$$(t, x, \delta) \mapsto x \{F_\beta(t - \beta'x) - \delta\} 1_{[\epsilon, 1-\epsilon]}(F_\beta(t - \beta'x))$$

to \mathcal{K}_β . We denote by $H(\zeta, \mathcal{K}_\beta, \mathbb{P}_n)$ the random ζ -entropy w.r.t. the L_2 -distance d_n , defined by

$$(S10.7) \quad d_n(k_1, k_2)^2 = \int \|k_1 - k_2\|^2 d\mathbb{P}_n, \quad k_1, k_2 \in \mathcal{K}_\beta.$$

Note that

$$x \{F_\beta(t - \beta'x) - \delta\} 1_{[\epsilon, 1-\epsilon]}(F(t - \beta'x)) = f_{1,\beta}(t, x, \delta) f_{2,\beta}(t, x, \delta),$$

where

$$f_{1,\beta}(t, x, \delta) = x \{F_\beta(t - \beta'x) - \delta\},$$

and

$$f_{2,\beta}(t, x, \delta) = 1_{[\epsilon, 1-\epsilon]}(F(t - \beta'x)).$$

Since t and x vary over a bounded region and, by (A4), F_β is of bounded variation, $f_{1,\beta}$ is of bounded variation. Moreover,

$$f_{2,\beta}(t, x, \delta) = 1_{[\epsilon, 1-\epsilon]}(F(t - \beta'x)) = 1_{[\epsilon, 1]}(F(t - \beta'x)) - 1_{(1-\epsilon, 1]}(F(t - \beta'x)).$$

Since F is monotone, we have:

$$(S10.8) \quad 1_{[\epsilon, 1]}(F(t - \beta'x)) - 1_{(1-\epsilon, 1]}(F(t - \beta'x)) = 1_{[a_{\epsilon, F}, M]}(t - \beta'x) - 1_{(b_{\epsilon, F}, M]}(t - \beta'x)$$

for points $a_{\epsilon, F} \leq b_{\epsilon, F}$, where M is an upper bound for the values of $t - \beta'x$. Hence $f_{2,\beta}$ is also a function of uniformly bounded variation.

We therefore get, using [1],

$$\sup_{\zeta > 0} \zeta H(\zeta, \mathcal{K}_\beta, \mathbb{P}_n) = O_p(1),$$

which implies:

$$\int_0^\zeta H(u, \mathcal{K}_\beta, \mathbb{P}_n)^{1/2} du = O_p(\zeta^{1/2}), \quad \zeta > 0.$$

This implies

$$\begin{aligned} & \int_{\hat{F}_{n,\beta}(t-\beta'x) \in [\epsilon, 1-\epsilon]} x \{F_\beta(t - \beta'x) - \delta\} d\mathbb{P}_n(t, x, \delta) \\ &= \int_{\hat{F}_{n,\beta}(t-\beta'x) \in [\epsilon, 1-\epsilon]} x \{F_\beta(t - \beta'x) - \delta\} dP_0(t, x, \delta) \\ & \quad + \int_{\hat{F}_{n,\beta}(t-\beta'x) \in [\epsilon, 1-\epsilon]} x \{F_\beta(t - \beta'x) - \delta\} d(\mathbb{P}_n - P_0)(t, x, \delta) \\ &= \int_{F_\beta(t-\beta'x) \in [\epsilon, 1-\epsilon]} x \{F_\beta(t - \beta'x) - \delta\} dP_0(t, x, \delta) \\ & \quad + \int_{F_\beta(t-\beta'x) \in [\epsilon, 1-\epsilon]} x \{F_\beta(t - \beta'x) - \delta\} d(\mathbb{P}_n - P_0)(t, x, \delta) + o_p(1) \\ &= \psi_{1,\epsilon}(\beta) + o_p(1), \end{aligned}$$

by the convergence in probability (and almost surely) of $\hat{F}_{n,\beta}$ to F_β , where we use Lemma S9.1 for the second term on the right-hand side of the first equality to make the transition of the integration region $\hat{F}_{n,\beta}(t-\beta'x) \in [\epsilon, 1-\epsilon]$ to $F_\beta(t-\beta'x) \in [\epsilon, 1-\epsilon]$. Here and in the sequel we use Lemma S9.1 without specifying the envelope or checking explicitly the permissibility which in our simple setting are rather trivial matters.

For the second and third term of (S10.5) we argue similarly, this time using the function class

$$(S10.9) \quad \mathcal{K}'_{\beta} = \{(t, x, \delta) \mapsto x\{F(t - \beta'x) - F_{\beta}(t - \beta'x)\}1_{[\epsilon, 1-\epsilon]}(F(t - \beta'x)) : F \in \mathcal{F}, \beta \in \Theta\}.$$

to which we add the function that is identically zero. This implies that these terms are $o_p(1)$.

Relation (S10.4) now follows. □

PROOF OF THEOREM 4.1, PART 2 (CONSISTENCY). By the definition of $\hat{\beta}_n$, we have

$$\int_{\hat{F}_{n,\beta}(t-\beta'x) \in [\epsilon, 1-\epsilon]} x\{\delta - \hat{F}_{n,\hat{\beta}_n}(t - \hat{\beta}'_n x)\} d\mathbb{P}_n(t, x, \delta) = 0.$$

We assume that $\hat{\beta}_n$ is contained in the compact set Θ , and hence the sequence $(\hat{\beta}_n)$ has a subsequence $(\hat{\beta}_{n_k} = \hat{\beta}_{n_k}(\omega))$, converging to an element β_* . If $\hat{\beta}_{n_k} = \hat{\beta}_{n_k}(\omega) \rightarrow \beta_*$, we get

$$\hat{F}_{n_k, \hat{\beta}_{n_k}}(t - \hat{\beta}'_{n_k} x) \rightarrow F_{\beta_*}(t - \beta'_* x),$$

where F_{β} is defined in (3.2). In the limit we get therefore the relation

$$\begin{aligned} & \lim_{k \rightarrow \infty} \int_{\hat{F}_{n_k, \hat{\beta}_{n_k}}(t - \hat{\beta}'_{n_k} x) \in [\epsilon, 1-\epsilon]} x\{\delta - F_{n_k, \hat{\beta}_{n_k}}(t - \hat{\beta}'_{n_k} x)\} d\mathbb{P}_{n_k}(t, x, \delta) \\ &= \int_{F_{\beta_*}(t - \beta'_* x) \in [\epsilon, 1-\epsilon]} x\{F_0(t - \beta'_0 x) - F_{\beta_*}(t - \beta'_* x)\} dG(t, x) = 0. \end{aligned}$$

Consider

$$\begin{aligned} & \int_{F_{\beta_*}(t - \beta'_* x) \in [\epsilon, 1-\epsilon]} x\{F_0(t - \beta'_0 x) - F_{\beta_*}(t - \beta'_* x)\} dG(t, x) \\ &= \int_{F_{\beta_*}(t - \beta'_* x) \in [\epsilon, 1-\epsilon]} x\{F_0(t - \beta'_* x + (\beta_* - \beta_0)'x) - F_{\beta_*}(t - \beta'_* x)\} dG(t, x). \end{aligned}$$

Since

$$F_{\beta_*}(t - \beta'_* x) = \int F_0(t - \beta'_* x + (\beta_* - \beta_0)'y) f_{X|T-\beta'_* X}(y|T - \beta'_* X = t - \beta'_* x) dy,$$

we get:

$$\begin{aligned}
& (\beta_* - \beta_0)' \int_{F_{\beta}(t - \beta'_* x) \in [\epsilon, 1 - \epsilon]} x \left\{ F_0(t - \beta'_* x + (\beta_* - \beta_0)' x) - F_{\beta_*}(t - \beta'_* x) \right\} dG(t, x) \\
&= \int_{F_{\beta_*}(t - \beta'_* x) \in [\epsilon, 1 - \epsilon]} (\beta_* - \beta_0)' x \left\{ F_0(t - \beta'_* x + (\beta_* - \beta_0)' x) \right. \\
&\quad \left. - \int F_0(t - \beta'_* x + (\beta_* - \beta_0)' y) f_{X|T - \beta'_* X}(y | T - \beta'_* X = t - \beta'_* x) dy \right\} dG(t, x) \\
&= \int_{F_{\beta_*}(u) \in [\epsilon, 1 - \epsilon]} \text{cov} \left\{ (\beta_* - \beta_0)' X, F_0(u + (\beta_* - \beta_0)' X) \mid T - \beta'_* X = u \right\} f_{T - \beta'_* X}(u) du \\
&= 0.
\end{aligned}$$

Suppose that $\beta_* \neq \beta_0$, then this integral can only be zero if $\text{cov}((\beta_* - \beta_0)' X, F_0(u + (\beta_* - \beta_0)' X) | T - \beta'_* X = u)$ is zero for all u such that $F_{\beta_*}(u) \in [\epsilon, 1 - \epsilon]$, if $f_{T - \beta'_* X}(u)$ stays away from zero on this region (Assumptions (A3)), using continuity of the functions in the integrand (Assumptions (A5)) and the non negativity of the conditional covariance function (see also Remark 4.2). Since this is excluded by the condition that the covariance $\text{covar}(X, F_0(u + (\beta - \beta_0)' X) | T - \beta' X = u)$ is continuous in u and not identically zero for u in the region $\{u : \epsilon \leq F_{\beta}(u) \leq 1 - \epsilon\}$, for each $\beta \in \Theta$, we must have: $\beta_* = \beta_0$. \square

PROOF OF THEOREM 4.1, PART 3 (ASYMPTOTIC NORMALITY). Before working out the details, we give a kind of “road map” for the proof of Theorem 4.1, Part 3.

1. We start with the relation

$$\psi_{1,n}^{(\epsilon)}(\hat{\beta}_n) = 0,$$

see (4.2), and show:

$$\begin{aligned}
& \psi_{1,n}^{(\epsilon)}(\hat{\beta}_n) \\
&= \int_{F_0(t - \beta'_0 x) \in [\epsilon, 1 - \epsilon]} \left\{ x - \phi_0(t - \beta'_0 x) \right\} \left\{ F_{\hat{\beta}_n}(t - \hat{\beta}'_n x) - F_0(t - \beta'_0 x) \right\} dP_0(t, x, \delta) \\
&\quad + \int_{F_0(t - \beta'_0 x) \in [\epsilon, 1 - \epsilon]} \left\{ x - \phi_0(t - \beta'_0 x) \right\} \left\{ F_0(t - \beta'_0 x) - \delta \right\} d(\mathbb{P}_n - P_0)(t, x, \delta) \\
&\quad (S10.10) \\
&\quad + o_p \left(n^{-1/2} + \hat{\beta}_n - \beta_0 \right),
\end{aligned}$$

where

$$\phi_0(u) = \phi_{\beta_0}(u),$$

and where ϕ_β is defined by:

$$(S10.11) \quad \phi_\beta(u) = \mathbb{E} \{X|T - \beta'X = u\}.$$

Since $\hat{\beta}_n \xrightarrow{p} \beta_0$ and

$$\begin{aligned} & \int_{F_0(t-\beta'_0x) \in [\epsilon, 1-\epsilon]} \left\{x - \phi_0(t - \beta'_0x)\right\} \left\{F_{\hat{\beta}_n}(t - \hat{\beta}'_n x) - F_0(t - \beta'_0x)\right\} dP_0(t, x, \delta) \\ &= \psi'_{1,\epsilon}(\beta_0) (\hat{\beta}_n - \beta_0) + o_p(\hat{\beta}_n - \beta_0), \end{aligned}$$

this yields, using the invertibility of $\psi'_{1,\epsilon}(\beta_0)$,

$$\begin{aligned} & \sqrt{n}(\hat{\beta}_n - \beta_0) \\ &= -\psi'_{1,\epsilon}(\beta_0)^{-1} \left\{ \sqrt{n} \int_{F_0(t-\beta'_0x) \in [\epsilon, 1-\epsilon]} \left\{x - \phi_{\beta_0}(t - \beta'_0x)\right\} \left\{F_0(t - \beta'_0x) - \delta\right\} \right. \\ & \quad \left. d(\mathbb{P}_n - P_0)(t, x, \delta) \right\} + o_p\left(1 + \sqrt{n}(\hat{\beta}_n - \beta_0)\right). \end{aligned}$$

As a consequence, the result of Theorem 4.1 follows, since

$$\begin{aligned} & \sqrt{n} \int_{F_0(t-\beta'_0x) \in [\epsilon, 1-\epsilon]} \left\{x - \phi_0(t - \beta'_0x)\right\} \left\{F_0(t - \beta'_0x) - \delta\right\} d(\mathbb{P}_n - P_0)(t, x, \delta) \\ & \xrightarrow{d} N(0, B). \end{aligned}$$

2. To show that (S10.10) holds, we need entropy results for the functions $u \mapsto \hat{F}_{n,\beta}(u)$ and $u \mapsto \bar{\phi}_{\beta, \hat{F}_{n,\beta}}(u)$ (see (S10.12) below). We also have to deal with the simpler parametric functions F_β and ϕ_β , parametrized by the finite dimensional parameter β , which are the population equivalents of $\hat{F}_{n,\beta}$ and $\bar{\phi}_{\beta, \hat{F}_{n,\beta}}$.
3. The result will then follow from the properties of F_β and ϕ_β , together with the closeness of $\hat{F}_{n,\beta}$ to F_β and $\bar{\phi}_{\beta, \hat{F}_{n,\beta}}$ to ϕ_β , respectively, and the convergence of $\hat{\beta}_n$ to β_0 .

Let $\bar{\phi}_{\hat{\beta}_n, \hat{F}_{n, \hat{\beta}_n}}$ be a (random) piecewise constant version of $\phi_{\hat{\beta}_n}$, where, for a piecewise constant distribution function F with finitely many jumps at $\tau_1 < \tau_2 < \dots$, the function $\bar{\phi}_{\beta, F}$ is defined in the following way.

$$(S10.12) \quad \bar{\phi}_{\beta, F}(u) = \begin{cases} \phi_\beta(\tau_i), & \text{if } F_\beta(u) > F(\tau_i), u \in [\tau_i, \tau_{i+1}), \\ \phi_\beta(s), & \text{if } F_\beta(u) = F(s), \text{ for some } s \in [\tau_i, \tau_{i+1}), \\ \phi_\beta(\tau_{i+1}), & \text{if } F_\beta(u) < F(\tau_i), u \in [\tau_i, \tau_{i+1}). \end{cases}$$

We can write:

$$\begin{aligned}
\psi_{1,n}^{(\epsilon)}(\hat{\beta}_n) &= \int_{\hat{F}_{n,\hat{\beta}_n}(t-\hat{\beta}'_n x) \in [\epsilon, 1-\epsilon]} x \{ \hat{F}_{n,\hat{\beta}_n}(t-\hat{\beta}'_n x) - \delta \} d\mathbb{P}_n(t, x, \delta) \\
&= \int_{\hat{F}_{n,\hat{\beta}_n}(t-\hat{\beta}'_n x) \in [\epsilon, 1-\epsilon]} \{ x - \phi_{\hat{\beta}_n}(t-\hat{\beta}'_n x) \} \{ \hat{F}_{n,\hat{\beta}_n}(t-\hat{\beta}'_n x) - \delta \} d\mathbb{P}_n(t, x, \delta) \\
&\quad + \int_{\hat{F}_{n,\hat{\beta}_n}(t-\hat{\beta}'_n x) \in [\epsilon, 1-\epsilon]} \{ \phi_{\hat{\beta}_n}(t-\hat{\beta}'_n x) - \bar{\phi}_{\hat{\beta}_n, \hat{F}_{n,\hat{\beta}_n}}(t-\hat{\beta}'_n x) \} \\
&\quad \cdot \{ \hat{F}_{n,\hat{\beta}_n}(t-\hat{\beta}'_n x) - \delta \} d\mathbb{P}_n(t, x, \delta) \\
&= I + II,
\end{aligned}$$

using

$$\int_{\hat{F}_{n,\hat{\beta}_n}(t-\hat{\beta}'_n x) \in [\epsilon, 1-\epsilon]} \bar{\phi}_{\hat{\beta}_n, \hat{F}_{n,\hat{\beta}_n}}(t-\hat{\beta}'_n x) \{ \hat{F}_{n,\hat{\beta}_n}(t-\hat{\beta}'_n x) - \delta \} d\mathbb{P}_n(t, x, \delta) = 0,$$

by the definition of the MLE $\hat{F}_{n,\hat{\beta}_n}$ as the slope of the greatest convex minorant of the corresponding cusum diagram, based on the values of the Δ_i in the ordering of the $T_i - \hat{\beta}'_n X_i$ (see also Lemma A.5 on p.380 of [13]). Since the function $u \mapsto \phi_{\hat{\beta}_n}(u)$ has a totally bounded derivative (as a consequence of (S10.11) and assumption (A5)), we can bound the Euclidean norm of the differences $\phi_{\hat{\beta}_n}(u) - \bar{\phi}_{\hat{\beta}_n, \hat{F}_{n,\hat{\beta}_n}}(u)$ above by a constant times $|\hat{F}_{n,\hat{\beta}_n}(u) - F_{\hat{\beta}_n}(u)|$, for $u \in A_{\epsilon,\beta}$ (see (A2)), i.e.,

$$\text{(S10.13)} \quad \|\phi_{\hat{\beta}_n}(u) - \bar{\phi}_{\hat{\beta}_n, \hat{F}_{n,\hat{\beta}_n}}(u)\| \leq K |\hat{F}_{n,\hat{\beta}_n}(u) - F_{\hat{\beta}_n}(u)|,$$

for some constant $K > 0$ (see for this technique for example (10.64) in [14]). Note that we need $f_{\hat{\beta}_n}(u) > 0$ for applying this, which is ensured by (A2). Also note that a component of $\bar{\phi}_{\hat{\beta}_n, \hat{F}_{n,\hat{\beta}_n}}$ is a uniformly totally bounded function, since the total variation is bounded by the total variation of the corresponding component of $u \mapsto \phi_{\hat{\beta}_n}(u)$ which is finite by (A5).

We have:

$$\begin{aligned}
II &= \int_{\hat{F}_{n,\hat{\beta}_n}(t-\hat{\beta}'_n x) \in [\epsilon, 1-\epsilon]} \left\{ \phi_{\hat{\beta}_n}(t - \hat{\beta}'_n x) - \bar{\phi}_{\hat{\beta}_n, \hat{F}_{n,\hat{\beta}_n}}(t - \hat{\beta}'_n x) \right\} \\
&\quad \cdot \left\{ \hat{F}_{n,\hat{\beta}_n}(t - \hat{\beta}'_n x) - \delta \right\} d\mathbb{P}_n(t, x, \delta) \\
&= \int_{\hat{F}_{n,\hat{\beta}_n}(t-\hat{\beta}'_n x) \in [\epsilon, 1-\epsilon]} \left\{ \phi_{\hat{\beta}_n}(t - \hat{\beta}'_n x) - \bar{\phi}_{\hat{\beta}_n, \hat{F}_{n,\hat{\beta}_n}}(t - \hat{\beta}'_n x) \right\} \\
&\quad \cdot \left\{ \hat{F}_{n,\hat{\beta}_n}(t - \hat{\beta}'_n x) - \delta \right\} d(\mathbb{P}_n - P_0)(t, x, \delta) \\
&\quad + \int_{\hat{F}_{n,\hat{\beta}_n}(u) \in [\epsilon, 1-\epsilon]} \left\{ \phi_{\hat{\beta}_n}(u) - \bar{\phi}_{\hat{\beta}_n, \hat{F}_{n,\hat{\beta}_n}}(u) \right\} \left\{ \hat{F}_{n,\hat{\beta}_n}(u) - F_{\hat{\beta}_n}(u) \right\} f_{T-\hat{\beta}'_n X}(u) du \\
&\quad + \int_{\hat{F}_{n,\hat{\beta}_n}(t-\hat{\beta}'_n x) \in [\epsilon, 1-\epsilon]} \left\{ \phi_{\hat{\beta}_n}(t - \hat{\beta}'_n x) - \bar{\phi}_{\hat{\beta}_n, \hat{F}_{n,\hat{\beta}_n}}(t - \hat{\beta}'_n x) \right\} \\
&\quad \cdot \left\{ F_{\hat{\beta}_n}(t - \hat{\beta}'_n x) - F_0(t - \beta'_0 x) \right\} dP_0(t, x, \delta) \\
&= II_a + II_b + II_c.
\end{aligned}$$

First consider II_a . For this term we use an argument similar to the arguments used in [10] (for similar purposes, see, e.g. Lemma 3 in [10]) and Lemma 10.11, p. 308 in [14]. Let \mathcal{F} be the set of piecewise constant distribution functions with finitely many jumps (like the MLE $\hat{F}_{n,\hat{\beta}_n}$), and let \mathcal{K}_1 be the set of functions

$$\begin{aligned}
\mathcal{K}_1 &= \left\{ (t, x, \delta) \mapsto (\phi_{\beta}(t - \beta'x) - \bar{\phi}_{\beta, F}(t - \beta'x))(F(t - \beta'x) - \delta) \right. \\
&\quad \cdot \mathbf{1}_{[\epsilon, 1-\epsilon]}(F(t - \beta'x)) : F \in \mathcal{F}, \beta \in \Theta \left. \right\},
\end{aligned}
\tag{S10.14}$$

where $\bar{\phi}_{\beta, F}$ is again defined by (S10.12). We add the function which is identically zero to \mathcal{K}_1 .

The functions $u \mapsto F(u)$, for $F \in \mathcal{F}$ and (as argued above) $u \mapsto \bar{\phi}_{\beta, F}(u)$ are bounded functions of uniformly bounded variation. We denote by $H(\zeta, \mathcal{K}_1, \mathbb{P}_n)$ the random ζ -entropy w.r.t. the L_2 -distance d_n , defined by

$$d_n(k_1, k_2)^2 = \int \|k_1 - k_2\|^2 d\mathbb{P}_n, \quad k_1, k_2 \in \mathcal{K}_1.
\tag{S10.15}$$

Note that, for $F_1, F_2 \in \mathcal{F}$,

$$\begin{aligned}
&F_1(t - \beta'_1 x) - F_2(t - \beta'_2 x) \\
&= F_1(t - \beta'_1 x) - F_{\beta_1}(t - \beta'_1 x) + F_{\beta_1}(t - \beta'_1 x) - F_{\beta_2}(t - \beta'_2 x) \\
&\quad + F_{\beta_2}(t - \beta'_2 x) - F_2(t - \beta'_2 x),
\end{aligned}$$

and that (see (3.2)):

$$\begin{aligned}
& |F_{\beta_1}(t - \beta'_1 x) - F_{\beta_2}(t - \beta'_2 x)| \\
&= \left| \int F_0(t - \beta'_0 x + (\beta_1 - \beta_0)'(y - x)) f_{X|T-\beta'_1 X}(y|t - \beta'_1 x) dy \right. \\
&\quad \left. - \int F_0(t - \beta'_0 x + (\beta_2 - \beta_0)'(y - x)) f_{X|T-\beta'_2 X}(y|t - \beta'_2 x) dy \right| \\
&= O(|\beta_1 - \beta_2|),
\end{aligned}$$

by (A2) and (A5).

For the indicator function $1_{[\epsilon, 1]}(F(t - \beta'x))$ we get, as in (S10.8), using the monotonicity of F ,

$$\begin{aligned}
& 1_{[\epsilon, 1]}(F(t - \beta'x)) \\
&= 1_{[\epsilon, 1]}(F(t - \beta'x)) - 1_{(1-\epsilon, 1]}(F(t - \beta'x)) = 1_{[a_{\epsilon, F}, M]}(t - \beta'x) - 1_{(b_{\epsilon, F}, M]}(t - \beta'x),
\end{aligned}$$

for points $a_{\epsilon, F} \leq b_{\epsilon, F}$, where M is an upper bound for the values of $t - \beta'x$, implying that the function

$$(t, x) \mapsto 1_{[\epsilon, 1]}(F(t - \beta'x)),$$

is of uniformly bounded variation. So the functions in \mathcal{K}_1 are products of functions of uniformly bounded variation, and we therefore get, using [1],

$$\sup_{\zeta > 0} \zeta H(\zeta, \mathcal{K}_1, \mathbb{P}_n) = O_p(1),$$

which implies:

$$\int_0^\zeta H(u, \mathcal{K}_1, \mathbb{P}_n)^{1/2} du = O_p(\zeta^{1/2}), \quad \zeta > 0.$$

Defining

$$\begin{aligned}
k_{\beta, F}(t, x, \delta) = & (\phi_\beta(t - \beta'x) - \bar{\phi}_{\beta, F}(t - \beta'x))(F(t - \beta'x) - \delta) \\
& \cdot 1_{[\epsilon, 1-\epsilon]}(F(t - \beta'x))
\end{aligned}$$

for $F \in \mathcal{F}$, we get, using (S10.13),

$$\begin{aligned}
& \left\{ \int \left\| k_{\hat{\beta}_n, \hat{F}_n, \hat{\beta}_n}(t, x, \delta) \right\|^2 dP_0(t, x, \delta) \right\}^2 \\
& \leq \int_{\hat{F}_n, \hat{\beta}_n} \left\| \phi_{\hat{\beta}_n}(t - \hat{\beta}'_n x) - \bar{\phi}_{\hat{\beta}_n, \hat{F}_n, \hat{\beta}_n}(t - \hat{\beta}'_n x) \right\|^2 dP_0(t, x, \delta) \\
& \leq K \int_{\hat{F}_n, \hat{\beta}_n} \left\{ \hat{F}_{n, \hat{\beta}_n}(t - \hat{\beta}'_n x) - F_{\hat{\beta}_n}(t - \hat{\beta}'_n x) \right\}^2 dP_0(t, x, \delta) \\
& \leq K' \int_{\hat{F}_n, \hat{\beta}_n} \left\{ \hat{F}_{n, \hat{\beta}_n}(u) - F_{\hat{\beta}_n}(u) \right\}^2 du \\
& \xrightarrow{p} 0,
\end{aligned}$$

for constants $K, K' > 0$. This implies

$$(S10.16) \quad \sqrt{n} II_a = \sqrt{n} \int k_{\hat{\beta}_n, \hat{F}_n, \hat{\beta}_n}(t, x, \delta) d(\mathbb{P}_n - P_0)(t, x, \delta) = o_p(1),$$

by an application of Lemma S9.1.

Using (S10.13), $\|F_{\hat{\beta}_n} - \hat{F}_{n, \hat{\beta}_n}\|_2 = O_p(n^{-1/3})$ and the Cauchy-Schwarz inequality on the second term we get,

$$II_b = O_p(n^{-2/3}).$$

The functions ϕ_β and F_β are of a simple parametric nature, since

$$\phi_\beta = \mathbb{E}(X|T - \beta'X),$$

and

$$F_\beta(u) = \int F_0(u + (\beta - \beta_0)'x) f_{X|T-\beta'X}(x|T - \beta'X = u) dx,$$

see (3.2). Moreover, since:

$$\begin{aligned}
F_{\hat{\beta}_n}(u) &= F_0(u) + (\hat{\beta}_n - \beta_0)' \int x f_0(u) f_{X|T-\hat{\beta}'_n X}(x|u) dx + o_p(\hat{\beta}_n - \beta_0) \\
(S10.17) \quad &= F_0(u) + (\hat{\beta}_n - \beta_0)' f_0(u) \mathbb{E}\{X|T - \hat{\beta}'_n X = u\} + o_p(\hat{\beta}_n - \beta_0),
\end{aligned}$$

and since the difference $\phi_{\hat{\beta}_n} - \bar{\phi}_{\hat{\beta}_n, \hat{F}_n, \hat{\beta}_n}$ converges to zero, we get for the third term II_c :

$$\begin{aligned}
II_c &= \int_{\hat{F}_n, \hat{\beta}_n} \left\{ \phi_{\hat{\beta}_n}(t - \hat{\beta}'_n x) - \bar{\phi}_{\hat{\beta}_n, \hat{F}_n, \hat{\beta}_n}(t - \hat{\beta}'_n x) \right\} \\
&\quad \cdot \left\{ F_{\hat{\beta}_n}(t - \hat{\beta}'_n x) - F_0(t - \beta'_0 x) \right\} dP_0(t, x, \delta) \\
&= o_p(\hat{\beta}_n - \beta_0).
\end{aligned}$$

We therefore conclude,

$$\psi_{1,n}^{(\epsilon)}(\hat{\beta}_n) = I + o_p\left(n^{-1/2} + \hat{\beta}_n - \beta_0\right).$$

We now write,

$$\begin{aligned} I &= \int_{\hat{F}_{n,\hat{\beta}_n}(t-\hat{\beta}'_n x) \in [\epsilon, 1-\epsilon]} \left\{x - \phi_{\hat{\beta}_n}(t - \hat{\beta}'_n x)\right\} \left\{\hat{F}_{n,\hat{\beta}_n}(t - \hat{\beta}'_n x) - \delta\right\} d\mathbb{P}_n(t, x, \delta) \\ &= \int_{\hat{F}_{n,\hat{\beta}_n}(t-\hat{\beta}'_n x) \in [\epsilon, 1-\epsilon]} \left\{x - \phi_{\hat{\beta}_n}(t - \hat{\beta}'_n x)\right\} \left\{F_{\hat{\beta}_n}(t - \hat{\beta}'_n x) - \delta\right\} d\mathbb{P}_n(t, x, \delta) \\ &\quad + \int_{\hat{F}_{n,\hat{\beta}_n}(t-\hat{\beta}'_n x) \in [\epsilon, 1-\epsilon]} \left\{x - \phi_{\hat{\beta}_n}(t - \hat{\beta}'_n x)\right\} \\ &\quad \cdot \left\{\hat{F}_{n,\hat{\beta}_n}(t - \hat{\beta}'_n x) - F_{\hat{\beta}_n}(t - \hat{\beta}'_n x)\right\} d\mathbb{P}_n(t, x, \delta) \\ &= I_a + I_b. \end{aligned}$$

We get:

$$\begin{aligned} I_a &= \int_{\hat{F}_{n,\hat{\beta}_n}(t-\hat{\beta}'_n x) \in [\epsilon, 1-\epsilon]} \left\{x - \phi_{\hat{\beta}_n}(t - \hat{\beta}'_n x)\right\} \left\{F_{\hat{\beta}_n}(t - \hat{\beta}'_n x) - \delta\right\} d\mathbb{P}_n(t, x, \delta) \\ &= \int_{\hat{F}_{n,\hat{\beta}_n}(t-\hat{\beta}'_n x) \in [\epsilon, 1-\epsilon]} \left\{x - \phi_{\hat{\beta}_n}(t - \hat{\beta}'_n x)\right\} \left\{F_{\hat{\beta}_n}(t - \hat{\beta}'_n x) - \delta\right\} d(\mathbb{P}_n - P_0)(t, x, \delta) \\ &\quad + \int_{\hat{F}_{n,\hat{\beta}_n}(t-\hat{\beta}'_n x) \in [\epsilon, 1-\epsilon]} \left\{x - \phi_{\hat{\beta}_n}(t - \hat{\beta}'_n x)\right\} \left\{F_{\hat{\beta}_n}(t - \hat{\beta}'_n x) - F_0(t - \beta'_0 x)\right\} dP_0(t, x, \delta). \end{aligned}$$

For the second integral on the right-hand side we get:

$$\begin{aligned} &\int_{\hat{F}_{n,\hat{\beta}_n}(t-\hat{\beta}'_n x) \in [\epsilon, 1-\epsilon]} \left\{x - \phi_{\hat{\beta}_n}(t - \hat{\beta}'_n x)\right\} \left\{F_{\hat{\beta}_n}(t - \hat{\beta}'_n x) - F_0(t - \beta'_0 x)\right\} dP_0(t, x, \delta) \\ &= \int_{\hat{F}_{n,\hat{\beta}_n}(t-\hat{\beta}'_n x) \in [\epsilon, 1-\epsilon]} \left\{x - \phi_{\hat{\beta}_n}(t - \hat{\beta}'_n x)\right\} \left\{F_{\hat{\beta}_n}(t - \hat{\beta}'_n x) - F_0(t - \hat{\beta}'_n x)\right\} dP_0(t, x, \delta) \\ &\quad + \int_{\hat{F}_{n,\hat{\beta}_n}(t-\hat{\beta}'_n x) \in [\epsilon, 1-\epsilon]} \left\{x - \phi_{\hat{\beta}_n}(t - \hat{\beta}'_n x)\right\} \left\{F_0(t - \hat{\beta}'_n x) - F_0(t - \beta'_0 x)\right\} dP_0(t, x, \delta), \end{aligned}$$

and next we get, using the definition of ϕ_β given in (S10.11), for the first integral on

the right-hand side of the last display:

$$\begin{aligned}
& \int_{\hat{F}_{n,\hat{\beta}_n}(t-\hat{\beta}'_n x) \in [\epsilon, 1-\epsilon]} \left\{ x - \phi_{\hat{\beta}_n}(t - \hat{\beta}'_n x) \right\} \left\{ F_{\hat{\beta}_n}(t - \hat{\beta}'_n x) - F_0(t - \hat{\beta}'_n x) \right\} dP_0(t, x, \delta) \\
&= \int_{\hat{F}_{n,\hat{\beta}_n}(u) \in [\epsilon, 1-\epsilon]} \left\{ x - \phi_{\hat{\beta}_n}(u) \right\} \left\{ F_{\hat{\beta}_n}(u) - F_0(u) \right\} f_{T-\hat{\beta}_n X}(u) f_{X|T-\hat{\beta}_n X}(x|u) du dx \\
&= 0.
\end{aligned}$$

Furthermore, we get by expanding $F_0(t - \beta'x)$ and by the continuity of $\beta \mapsto \phi_\beta(u)$ at $\beta = \beta_0$

$$\begin{aligned}
& \int_{\hat{F}_{n,\hat{\beta}_n}(t-\hat{\beta}'_n x) \in [\epsilon, 1-\epsilon]} \left\{ x - \phi_{\hat{\beta}_n}(t - \hat{\beta}'_n x) \right\} \left\{ F_0(t - \hat{\beta}'_n x) - F_0(t - \beta'_0 x) \right\} dP_0(t, x, \delta) \\
&= \int_{\hat{F}_{n,\hat{\beta}_n}(t-\hat{\beta}'_n x) \in [\epsilon, 1-\epsilon]} \left\{ x - \phi_{\hat{\beta}_n}(t - \hat{\beta}'_n x) \right\} (\hat{\beta}_n - \beta_0)' x f_0(t - \beta'_0 x) dP_0(t, x, \delta) \\
&\quad + o_p(\hat{\beta}_n - \beta_0) \\
&= \left\{ \int_{\hat{F}_{n,\hat{\beta}_n}(t-\hat{\beta}'_n x) \in [\epsilon, 1-\epsilon]} \left\{ x - \phi_0(t - \beta'_0 x) \right\} x' f_0(t - \beta'_0 x) dP_0(t, x, \delta) \right\} (\hat{\beta}_n - \beta_0) \\
&\quad + o_p(\hat{\beta}_n - \beta_0).
\end{aligned}$$

Finally we get from the consistency of $\hat{F}_{n,\hat{\beta}_n}$:

$$\begin{aligned}
& \left\{ \int_{\hat{F}_{n,\hat{\beta}_n}(t-\hat{\beta}'_n x) \in [\epsilon, 1-\epsilon]} \left\{ x - \phi_0(t - \beta'_0 x) \right\} x' f_0(t - \beta'_0 x) dP_0(t, x, \delta) \right\} (\hat{\beta}_n - \beta_0) \\
&= \left\{ \int_{F_0(t-\beta'_0 x) \in [\epsilon, 1-\epsilon]} \left\{ x - \phi_0(t - \beta'_0 x) \right\} x' f_0(t - \beta'_0 x) dP_0(t, x, \delta) \right\} (\hat{\beta}_n - \beta_0) \\
&\quad + o_p(\hat{\beta}_n - \beta_0) \\
&= \psi'_{1,\epsilon}(\beta_0)(\hat{\beta}_n - \beta_0) + o_p(\hat{\beta}_n - \beta_0).
\end{aligned}$$

So we obtain:

$$\begin{aligned}
I_a &= \int_{\hat{F}_{n,\hat{\beta}_n}(t-\hat{\beta}'_n x) \in [\epsilon, 1-\epsilon]} \left\{ x - \phi_{\hat{\beta}_n}(t - \hat{\beta}'_n x) \right\} \left\{ F_{\hat{\beta}_n}(t - \hat{\beta}'_n x) - \delta \right\} d(\mathbb{P}_n - P_0)(t, x, \delta) \\
&\quad + \psi'_{1,\epsilon}(\beta_0)(\hat{\beta}_n - \beta_0) + o_p(\hat{\beta}_n - \beta_0).
\end{aligned}$$

We now proceed again as before, and define \mathcal{K}'_1 to be the set of functions

$$\mathcal{K}'_1 = \left\{ (t, x, \delta) \mapsto (x - \phi_\beta(t - \beta'x)) (F_\beta(t - \beta'x) - \delta) 1_{[\epsilon, 1-\epsilon]}(F(t - \beta'x)) \right. \\ \left. : F \in \mathcal{F}, \beta \in \Theta \right\}.$$

We add the function

$$(t, x, \delta) \mapsto (x - \phi_0(t - \beta'_0x)) (F_0(t - \beta'_0x) - \delta) 1_{[\epsilon, 1-\epsilon]}(F_0(t - \beta'_0x))$$

to the set \mathcal{K}'_1 . The pseudo-distance d_n is defined by (S10.15) again, with \mathcal{K}_1 replaced by \mathcal{K}'_1 . We therefore get, similarly as before, using [1],

$$\sup_{\zeta > 0} \zeta H(\zeta, \mathcal{K}'_1, \mathbb{P}_n) = O_p(1),$$

implying

$$\int_0^\zeta H(u, \mathcal{K}'_1, \mathbb{P}_n)^{1/2} du = O_p(\zeta^{1/2}), \quad \zeta > 0.$$

Moreover, we get:

$$\begin{aligned} & (x - \phi_{\hat{\beta}_n}(t - \hat{\beta}'_n x)) (F_{\hat{\beta}_n}(t - \hat{\beta}'_n x) - \delta) 1_{[\epsilon, 1-\epsilon]}(\hat{F}_{n, \hat{\beta}_n}(t - \hat{\beta}'_n x)) \\ & - (x - \phi_0(t - \beta'_0 x)) (F_0(t - \beta'_0 x) - \delta) 1_{[\epsilon, 1-\epsilon]}(F_0(t - \beta'_0 x)) \\ & = \left\{ (x - \phi_{\hat{\beta}_n}(t - \hat{\beta}'_n x)) (F_{\hat{\beta}_n}(t - \hat{\beta}'_n x) - \delta) \right. \\ & \quad \left. - (x - \phi_0(t - \beta'_0 x)) (F_0(t - \beta'_0 x) - \delta) \right\} 1_{[\epsilon, 1-\epsilon]}(\hat{F}_{n, \hat{\beta}_n}(t - \hat{\beta}'_n x)) \\ & \quad + (x - \phi_0(t - \beta'_0 x)) (F_0(t - \beta'_0 x) - \delta) \\ & \quad \cdot \left\{ 1_{[\epsilon, 1-\epsilon]}(F_0(t - \beta'_0 x)) - 1_{[\epsilon, 1-\epsilon]}(\hat{F}_{n, \hat{\beta}_n}(t - \hat{\beta}'_n x)) \right\} \\ & = A_n(t, x, \delta) + B_n(t, x, \delta), \end{aligned}$$

implying

$$\begin{aligned} & \int \left\{ (x - \phi_{\hat{\beta}_n}(t - \hat{\beta}'_n x)) (F_{\hat{\beta}_n}(t - \hat{\beta}'_n x) - \delta) 1_{[\epsilon, 1-\epsilon]}(\hat{F}_{n, \hat{\beta}_n}(t - \hat{\beta}'_n x)) \right. \\ & \quad \left. - (x - \phi_0(t - \beta'_0 x)) (F_0(t - \beta'_0 x) - \delta) 1_{[\epsilon, 1-\epsilon]}(F_0(t - \beta'_0 x)) \right\}^2 dP_0(t, x, \delta) \\ & \leq 2 \int \{A_n(t, x, \delta)^2 + B_n(t, x, \delta)^2\} dP_0(t, x, \delta) = o_p(1), \end{aligned}$$

since the integrals w.r.t. A_n^2 and B_n^2 tends to zero using the consistency of $\hat{\beta}_n$ and $\hat{F}_{n, \hat{\beta}_n}$.

Hence we get from Lemma S9.1:

$$I_a = \int_{F_0(t-\beta'_0 x) \in [\epsilon, 1-\epsilon]} \left\{ x - \phi_0(t - \beta'_0 x) \right\} \left\{ F_0(t - \beta'_0 x) - \delta \right\} d(\mathbb{P}_n - P_0)(t, x, \delta) \\ + \psi'_{1,\epsilon}(\beta_0)(\hat{\beta}_n - \beta_0) + o_p(\hat{\beta}_n - \beta_0) + o_p(n^{-1/2}).$$

This means that we get the conclusion

$$\begin{aligned} & \int_{F_0(t-\beta'_0 x) \in [\epsilon, 1-\epsilon]} \left\{ x - \phi_0(t - \beta'_0 x) \right\} \left\{ F_0(t - \beta'_0 x) - \delta \right\} d(\mathbb{P}_n - P_0)(t, x, \delta) \\ (S10.18) \quad &= -\psi'_{1,\epsilon}(\beta_0)(\hat{\beta}_n - \beta_0) + o_p(\hat{\beta}_n - \beta_0) + o_p(n^{-1/2}), \end{aligned}$$

if we can show that I_b is negligible.

Since, by definition of ϕ_β given in (S10.11),

$$\int_{\hat{F}_{n,\hat{\beta}_n}(t-\hat{\beta}'_n x) \in [\epsilon, 1-\epsilon]} \left\{ x - \phi_{\hat{\beta}_n}(t - \hat{\beta}'_n x) \right\} f_{X|T-\hat{\beta}'_n X}(x|t - \hat{\beta}'_n x) dx = 0,$$

we have

$$I_b = \int_{\hat{F}_{n,\hat{\beta}_n}(t-\hat{\beta}'_n x) \in [\epsilon, 1-\epsilon]} \left\{ x - \phi_{\hat{\beta}_n}(t - \hat{\beta}'_n x) \right\} \\ \cdot \left\{ \hat{F}_{n,\hat{\beta}_n}(t - \hat{\beta}'_n x) - F_{\hat{\beta}_n}(t - \hat{\beta}'_n x) \right\} d(\mathbb{P}_n - P_0)(t, x, \delta).$$

The negligibility of I_b now follows in the same way as (S10.16), using the parametric nature of the function ϕ_β and the entropy properties of the class of functions

$$u \mapsto \hat{F}_{n,\hat{\beta}_n}(u) - F_{\hat{\beta}_n}(u).$$

The conclusion now follows from (S10.18). □

REMARK S10.1. Note that the proof above yields the representation

$$\begin{aligned} & \hat{\beta}_n - \beta_0 \\ & \sim -n^{-1} \psi'_{1,\epsilon}(\beta_0)^{-1} \sum_{i=1}^n (X_i - E(X_i - E(X|T - \beta'_0 X))) \{ \Delta_i - F_0(T_i - \beta'_0 X_i) \}, \end{aligned}$$

where $\psi'_{1,\epsilon}(\beta_0)$ is given by (S10.2).

S11. Asymptotic behavior of the efficient estimate based on the MLE $\hat{F}_{n,\beta}$. In this section we prove the asymptotic efficiency of the score estimator defined in Section 4.2. The proof of existence of the root and the consistency proof for the score estimator is similar to the proof of existence and consistency of the first score estimator defined in Section 4.1, thus omitted. Before starting the proof of the asymptotic normality, we first prove, in Lemma S11.1 given below, some results on the properties of the density estimator $f_{nh,\beta}$ defined in (4.3) which are needed for the proof of Theorem 4.2.

S11.1. *Lemma on the L_2 -distance between $f_{nh,\beta}$ and f_β .*

LEMMA S11.1. *Let the distribution function F_β be three times continuously differentiable on the interior of the support of $f_\beta = F'_\beta$. Then, for $h \asymp n^{-1/7}$,*

$$\begin{aligned} & \int_{\hat{F}_{n,\beta}(t-\beta'x) \in [\epsilon, 1-\epsilon]} \{f_{nh,\beta}(t-\beta'x) - f_\beta(t-\beta'x)\}^2 dG(t, x) \\ &= O_p\left(h^{-2}n^{-2/3}\right) + O_p(h^4) = O_p\left(n^{-8/21}\right), \end{aligned}$$

and

$$\int_{\epsilon/2 \leq F_\beta(u) \leq 1-\epsilon/2} f'_{nh,\beta}(u)^2 du = O_p(1).$$

PROOF. Let $A_{n,\epsilon,\beta} = \{u : \hat{F}_{n,\beta}(u) \in [\epsilon, 1-\epsilon]\}$. We have,

$$\begin{aligned} & \int_{A_{n,\epsilon,\beta}} \{f_{nh,\beta}(u) - f_\beta(u)\}^2 dP_\beta(u) \\ &= \int_{A_{n,\epsilon,\beta}} \left(\int h^{-1} K(h^{-1}\{u-w\}) d\hat{F}_{n,\beta}(w) - f_\beta(u) \right)^2 dP_\beta(u) \end{aligned}$$

where P_β is the probability measure of $T - \beta'X$. Using integration by parts, and introducing the function F_β , the last expression can be bounded by 2 times

$$\begin{aligned} & \int_{A_{n,\epsilon,\beta}} \left(\int_{u-h}^{u+h} h^{-2} K'(h^{-1}\{u-w\}) (\hat{F}_{n,\beta}(w) - F_\beta(w)) dw \right)^2 dP_\beta(u) \\ (S11.1) \quad & + \int_{A_{n,\epsilon,\beta}} \left(\int_{u-h}^{u+h} h^{-2} K'(h^{-1}\{u-w\}) F_\beta(w) dw - f_\beta(u) \right)^2 dP_\beta(u) \end{aligned}$$

By applying the Cauchy-Schwarz inequality on the integrand in the first line, we get:

$$\begin{aligned}
& \left(\int_{u-h}^{u+h} h^{-2} K'(h^{-1}\{u-w\}) \{ \hat{F}_{n,\beta}(w) - F_\beta(w) \} dw \right)^2 \\
& \leq h^{-4} \int_{u-h}^{u+h} K'(h^{-1}\{u-w\})^2 dw \int_{u-h}^{u+h} \{ \hat{F}_{n,\beta}(w) - F_\beta(w) \}^2 dw \\
& = h^{-3} \int_{-1}^1 K'(w)^2 dw \int_{u-h}^{u+h} \{ \hat{F}_{n,\beta}(w) - F_\beta(w) \}^2 dw \\
& = C_1 h^{-3} \int_{u-h}^{u+h} \{ \hat{F}_{n,\beta}(w) - F_\beta(w) \}^2 dw,
\end{aligned}$$

for $C_1 = \int_{-1}^1 K'(w)^2 dw$. By (11.32) and (11.33) on p. 327 of [14] we have:

$$\mathbb{E} \{ \hat{F}_{n,\beta}(w) - F_\beta(w) \}^2 \leq C_2 n^{-2/3},$$

for some $C_2 > 0$ uniformly in w in the domain of integration. Hence, using Markov's inequality, we get:

$$\begin{aligned}
& \int_{A_{n,\epsilon,\beta}} \left(\int_{u-h}^{u+h} h^{-2} K'(h^{-1}\{u-w\}) \{ \hat{F}_{n,\beta}(w) - F_\beta(w) \} dw \right)^2 dP_\beta(u) \\
& = O_p(h^{-2} n^{-2/3}).
\end{aligned}$$

For the second term of (S11.1), we get:

$$\int h^{-2} K'(h^{-1}\{u-w\}) F_\beta(w) dw = \int K_h(u-w) f_\beta(w) dw,$$

and hence, by a classical bias computation:

$$\begin{aligned}
& \int_{A_{n,\epsilon,\beta}} \left(\int_{u-h}^{u+h} h^{-2} K'(h^{-1}\{u-w\}) F_\beta(w) dw - f_\beta(u) \right)^2 dP_\beta(u) \\
& \int_{A_{n,\epsilon,\beta}} \left(\int K_h(u-w) \{ f(w) - f(u) \} dw \right)^2 dP_\beta(u) \\
& = O(h^4).
\end{aligned}$$

since F_β is three times continuously differentiable.

We continue with the second part of Lemma S11.1. Using (11.32) and (11.33) on p. 327 of [14] again we have:

$$\mathbb{E} \int_{u-h}^{u+h} \{ \hat{F}_{n,\beta}(w) - F_\beta(w) \}^2 dw = O(h n^{-2/3}),$$

uniformly in $u \in A_{n,\epsilon,\beta}$.

We can write for $u \in A_{n,\epsilon,\beta}$, using integration by parts and the Cauchy-Schwarz inequality in the third line,

$$\begin{aligned}
f'_{nh,\beta}(u) &= \int h^{-3} K''(h^{-1}\{u-w\}) F_{n,\beta}(w) dw \\
&= \int h^{-3} K''(h^{-1}\{u-w\}) F_{\beta}(w) dw + \int h^{-3} K''(h^{-1}\{u-w\}) \{\hat{F}_{n,\beta}(w) - F_{\beta}(w)\} dw \\
&\leq \int h^{-1} K(h^{-1}\{u-w\}) f''_{\beta}(w) dw \\
&\quad + \left(\int_{u-h}^{u+h} h^{-6} K''(h^{-1}\{u-w\})^2 dw \int_{u-h}^{u+h} \{\hat{F}_{n,\beta}(w) - F_{\beta}(w)\}^2 dw \right)^{1/2} \\
&= O_p \left(1 + h^{-5/2+1/2} n^{-1/3} \right)
\end{aligned}$$

uniformly in $u \in A_{n,\epsilon/2,\beta}$ for all large n . If we choose $h \asymp n^{-1/7}$, then $h^{-2} n^{-1/3} \asymp n^{-1/21}$, and hence:

$$f'_{nh,\beta}(u) = O_p(1),$$

uniformly in u such that $\hat{F}_{n,\beta}(u) \in [\epsilon, 1-\epsilon]$ for all large n . This concludes the proof of Lemma [S11.1](#). \square

S11.2. Asymptotic normality of the efficient score estimator.

PROOF OF THEOREM [4.2](#) (ASYMPTOTIC NORMALITY). Since the proof is very similar to the proof of Theorem [4.1](#), we only give the main steps of the proof. We prove that:

(S11.2)

$$\begin{aligned}
&\psi_{2,nh}^{(\epsilon)}(\hat{\beta}_n) \\
&= \int_{F_0(t-\beta'_0 x) \in [\epsilon, 1-\epsilon]} \frac{\{x f_0(t-\beta'_0 x) - \varphi_{\beta_0}(t-\beta'_0 x)\} \{F_0(t-\beta'_0 x) - \delta\}}{F_0(t-\beta'_0 x) \{1 - F_0(t-\beta'_0 x)\}} d\mathbb{P}_n(t, x, \delta) \\
&\quad + \psi'_{2,\epsilon}(\beta_0)(\hat{\beta}_n - \beta_0) + o_p \left(n^{-1/2} + (\hat{\beta}_n - \beta_0) \right),
\end{aligned}$$

where φ_{β} is defined by

$$(S11.3) \quad \varphi_{\beta}(t - \beta'x) = E(X|T - \beta'X = t - \beta'x) f_{\beta}(t - \beta'x),$$

and $\psi_{2,\epsilon}$ is defined by,

(S11.4)

$$\begin{aligned} \psi_{2,\epsilon}(\beta) &= \int_{F_\beta(t-\beta'x) \in [\epsilon, 1-\epsilon]} \frac{\{xf_\beta(t-\beta'x) - \varphi_\beta(t-\beta'x)\} \{F_\beta(t-\beta'x) - \delta\}}{F_\beta(t-\beta'x)\{1 - F_\beta(t-\beta'x)\}} dP_0(t, x, \delta). \end{aligned}$$

Straightforward calculations show that,

$$\begin{aligned} \psi'_{2,\epsilon}(\beta_0) &= - \int_{F_0(t-\beta'_0x) \in [\epsilon, 1-\epsilon]} \frac{\{xf_0(t-\beta'_0x) - \varphi_{\beta_0}(t-\beta'_0x)\}^2}{F_0(t-\beta'_0x)\{1 - F_0(t-\beta'_0x)\}} dP_0(t, x, \delta) \\ &= -\mathbb{E}_\epsilon \left\{ \frac{f_0(T-\beta'_0X)^2 \{X - \mathbb{E}(X|T-\beta'_0X)\} \{X - \mathbb{E}(X|T-\beta'_0X)\}'}{F_0(T-\beta'_0X)\{1 - F_0(T-\beta'_0X)\}} \right\} \\ &= -I_\epsilon(\beta_0). \end{aligned}$$

(See also the derivation of the derivative ψ'_ϵ for the first score equation in the proof of Theorem 4.1, Part 1). Since

$$\begin{aligned} \sqrt{n} \int_{F_0(t-\beta'_0x) \in [\epsilon, 1-\epsilon]} \frac{\{xf_0(t-\beta'_0x) - \varphi_{\beta_0}(t-\beta'_0x)\} \{F_0(t-\beta'_0x) - \delta\}}{F_0(t-\beta'_0x)\{1 - F_0(t-\beta'_0x)\}} d\mathbb{P}_n(t, x, \delta) \\ \xrightarrow{d} N(0, I_\epsilon(\beta_0)), \end{aligned}$$

(S11.2) implies, using the non-singularity of $\psi'_{2,\epsilon}(\beta_0)$ and the consistency of $\hat{\beta}_n$,

$$\begin{aligned} \sqrt{n}(\hat{\beta}_n - \beta_0) &= -\psi'_{2,\epsilon}(\beta_0)^{-1} \left\{ \sqrt{n} \int_{F_0(t-\beta'_0x) \in [\epsilon, 1-\epsilon]} \frac{xf_0(t-\beta'_0x) - \varphi_{\beta_0}(t-\beta'_0x)}{F_0(t-\beta'_0x)\{1 - F_0(t-\beta'_0x)\}} \right. \\ &\quad \cdot \left. \{F_0(t-\beta'_0x) - \delta\} d\mathbb{P}_n(t, x, \delta) \right\} \\ &\quad + o_p(1 + \sqrt{n}(\hat{\beta}_n - \beta_0)) \\ &\xrightarrow{d} N(0, I_\epsilon(\beta_0)^{-1}). \end{aligned}$$

Let, analogously to the start of the proof of Theorem 4.1, $\bar{\varphi}_{\hat{\beta}_n, \hat{F}_n, \hat{\beta}_n}$ be a (random) piecewise constant version of $\varphi_{\hat{\beta}_n}$, where, for a piecewise constant distribution function F with finitely many jumps at $\tau_1 < \tau_2 < \dots$, the function $\bar{\varphi}_{\beta, F}$ is defined in the following way.

$$(S11.5) \quad \bar{\varphi}_{\beta, F}(u) = \begin{cases} \varphi_\beta(\tau_i), & \text{if } F_\beta(u) > F(\tau_i), u \in [\tau_i, \tau_{i+1}), \\ \varphi_\beta(s), & \text{if } F_\beta(u) = F(s), \text{ for some } s \in [\tau_i, \tau_{i+1}), \\ \varphi_\beta(\tau_{i+1}), & \text{if } F_\beta(u) < F(\tau_i), u \in [\tau_i, \tau_{i+1}). \end{cases}$$

We now have:

$$\begin{aligned}
& \psi_{2,nh}^{(\epsilon)}(\beta) \\
&= \int_{\hat{F}_{n,\hat{\beta}_n}(t-\hat{\beta}'_n x) \in [\epsilon, 1-\epsilon]} x f_{nh,\hat{\beta}_n}(t-\hat{\beta}'_n x) \frac{\hat{F}_{n,\hat{\beta}_n}(t-\hat{\beta}'_n x) - \delta}{\hat{F}_{n,\hat{\beta}_n}(t-\hat{\beta}'_n x) \{1 - \hat{F}_{n,\hat{\beta}_n}(t-\hat{\beta}'_n x)\}} d\mathbb{P}_n(t, x, \delta) \\
&= \int_{\hat{F}_{n,\hat{\beta}_n}(t-\hat{\beta}'_n x) \in [\epsilon, 1-\epsilon]} \left\{ x f_{nh,\hat{\beta}_n}(t-\hat{\beta}'_n x) - \varphi_{\hat{\beta}_n}(t-\hat{\beta}'_n x) \right\} \\
&\quad \cdot \frac{\hat{F}_{n,\hat{\beta}_n}(t-\hat{\beta}'_n x) - \delta}{\hat{F}_{n,\hat{\beta}_n}(t-\hat{\beta}'_n x) \{1 - \hat{F}_{n,\hat{\beta}_n}(t-\hat{\beta}'_n x)\}} d\mathbb{P}_n(t, x, \delta) \\
&\quad + \int_{\hat{F}_{n,\hat{\beta}_n}(t-\hat{\beta}'_n x) \in [\epsilon, 1-\epsilon]} \left\{ \varphi_{\hat{\beta}_n}(t-\hat{\beta}'_n x) - \bar{\varphi}_{n,\hat{F}_{n,\hat{\beta}_n}}(t-\hat{\beta}'_n x) \right\} \\
&\quad \cdot \frac{\hat{F}_{n,\hat{\beta}_n}(t-\hat{\beta}'_n x) - \delta}{\hat{F}_{n,\hat{\beta}_n}(t-\hat{\beta}'_n x) \{1 - \hat{F}_{n,\hat{\beta}_n}(t-\hat{\beta}'_n x)\}} d\mathbb{P}_n(t, x, \delta) \\
&= I + II.
\end{aligned}$$

Let \mathcal{F} be the set of piecewise constant distribution functions with finitely many jumps (like the MLE $\hat{F}_{n,\hat{\beta}_n}$), and let \mathcal{K}_2 be the set of functions

$$\begin{aligned}
\mathcal{K}_2 = & \left\{ (t, x, \delta) \mapsto \left\{ \varphi_\beta(t - \beta'x) - \bar{\varphi}_{\beta,F}(t - \beta'x) \right\} \frac{F(t - \beta'x) - \delta}{F(t - \beta'x) \{1 - F(t - \beta'x)\}} \right. \\
& \left. \cdot 1_{[\epsilon, 1-\epsilon]}(F(t - \beta'x)) : F \in \mathcal{F}, \beta \in \Theta \right\},
\end{aligned}
\tag{S11.6}$$

where $\bar{\varphi}_{\beta,F}$ is defined by (S11.5). We add the function which is identically zero to \mathcal{K}_2 . As in the proof of Theorem 4.1, the functions are uniformly bounded and also of uniformly bounded variation, using conditions (A4) and (A5). For k_1 and k_2 in \mathcal{K}_2 , we define

$$d_n(k_1, k_2)^2 = \int \|k_1 - k_2\|^2 d\mathbb{P}_n, \quad k_1, k_2 \in \mathcal{K}_2.
\tag{S11.7}$$

For this pseudo-distance we then get, using [1],

$$\sup_{\zeta > 0} \zeta H(\zeta, \mathcal{K}_2, \mathbb{P}_n) = O_p(1),$$

implying:

$$\int_0^\zeta H(u, \mathcal{K}_2, \mathbb{P}_n)^{1/2} du = O_p\left(\zeta^{1/2}\right), \quad \zeta > 0.$$

Note that the indicator function keeps $F(t - \beta'x)$ away from zero and 1, which is essential for getting the bounded variation property.

Following the same steps as in the proof of Theorem 4.1, we get:

$$II = o_p \left(n^{-1/2} + \hat{\beta}_n - \beta_0 \right).$$

We now write,

$$\begin{aligned} I &= \int_{\hat{F}_{n,\hat{\beta}_n}(t-\hat{\beta}'_n x) \in [\epsilon, 1-\epsilon]} \left\{ x f_{nh,\hat{\beta}_n}(t - \hat{\beta}'_n x) - \varphi_{\hat{\beta}_n}(t - \hat{\beta}'_n x) \right\} \\ &\quad \cdot \frac{\hat{F}_{n,\hat{\beta}_n}(t - \hat{\beta}'_n x) - \delta}{\hat{F}_{n,\hat{\beta}_n}(t - \hat{\beta}'_n x) \{1 - \hat{F}_{n,\hat{\beta}_n}(t - \hat{\beta}'_n x)\}} d\mathbb{P}_n(t, x, \delta) \\ &= \int_{\hat{F}_{n,\hat{\beta}_n}(t-\hat{\beta}'_n x) \in [\epsilon, 1-\epsilon]} \left\{ x f_{nh,\hat{\beta}_n}(t - \hat{\beta}'_n x) - \varphi_{\hat{\beta}_n}(t - \hat{\beta}'_n x) \right\} \\ &\quad \cdot \frac{F_{\hat{\beta}_n}(t - \hat{\beta}'_n x) - \delta}{\hat{F}_{n,\hat{\beta}_n}(t - \hat{\beta}'_n x) \{1 - \hat{F}_{n,\hat{\beta}_n}(t - \hat{\beta}'_n x)\}} d\mathbb{P}_n(t, x, \delta) \\ &\quad + \int_{\hat{F}_{n,\hat{\beta}_n}(t-\hat{\beta}'_n x) \in [\epsilon, 1-\epsilon]} \left\{ x f_{nh,\hat{\beta}_n}(t - \hat{\beta}'_n x) - \varphi_{\hat{\beta}_n}(t - \hat{\beta}'_n x) \right\} \\ &\quad \cdot \frac{\hat{F}_{n,\hat{\beta}_n}(t - \hat{\beta}'_n x) - F_{\hat{\beta}_n}(t - \hat{\beta}'_n x)}{\hat{F}_{n,\hat{\beta}_n}(t - \hat{\beta}'_n x) \{1 - \hat{F}_{n,\hat{\beta}_n}(t - \hat{\beta}'_n x)\}} d\mathbb{P}_n(t, x, \delta) \\ &= I_a + I_b. \end{aligned}$$

We now get, using Lemma S11.1 and $\|\hat{F}_{n,\hat{\beta}_n} - F_{\hat{\beta}_n}\|_2 = O_p(n^{-1/3})$,

$$I_b = O_p \left(n^{-11/21} \right) = o_p \left(n^{-1/2} \right).$$

Finally,

$$\begin{aligned} I_a &= \int_{\hat{F}_{n,\hat{\beta}_n}(t-\hat{\beta}'_n x) \in [\epsilon, 1-\epsilon]} \left\{ x f_{nh,\hat{\beta}_n}(t - \hat{\beta}'_n x) - \varphi_{\hat{\beta}_n}(t - \hat{\beta}'_n x) \right\} \\ &\quad \cdot \frac{\hat{F}_{n,\hat{\beta}_n}(t - \hat{\beta}'_n x) - \delta}{\hat{F}_{n,\hat{\beta}_n}(t - \hat{\beta}'_n x) \{1 - \hat{F}_{n,\hat{\beta}_n}(t - \hat{\beta}'_n x)\}} d(\mathbb{P}_n - P_0)(t, x, \delta) \\ &\quad + \int_{\hat{F}_{n,\hat{\beta}_n}(t-\hat{\beta}'_n x) \in [\epsilon, 1-\epsilon]} \left\{ x f_{nh,\hat{\beta}_n}(t - \hat{\beta}'_n x) - \varphi_{\hat{\beta}_n}(t - \hat{\beta}'_n x) \right\} \\ &\quad \cdot \frac{\hat{F}_{n,\hat{\beta}_n}(t - \hat{\beta}'_n x) - F_0(t - \beta'_0 x)}{\hat{F}_{n,\hat{\beta}_n}(t - \hat{\beta}'_n x) \{1 - \hat{F}_{n,\hat{\beta}_n}(t - \hat{\beta}'_n x)\}} dP_0(t, x, \delta). \end{aligned}$$

This time we consider the class of functions

$$\mathcal{K}'_2 = \left\{ (t, x, \delta) \mapsto (xf(t - \beta'x) - \varphi_\beta(t - \beta'x)) \frac{F(t - \beta'x) - \delta}{F(t - \beta'x)\{1 - F(t - \beta'x)\}} 1_{[\epsilon, 1-\epsilon]}(F(t - \beta'x)) \right. \\ \left. : F \in \mathcal{F}, f \in \mathcal{F}', \beta \in \Theta \right\},$$

where \mathcal{F}' is a class of uniformly bounded functions of uniformly bounded variation (which have the interpretation of estimates of F'_β), to which we add the function

$$(t, x, \delta) \mapsto (xf_0(t - \beta'_0x) - \varphi_{\beta_0}(t - \beta'_0x)) \frac{F_0(t - \beta'_0x) - \delta}{F_0(t - \beta'_0x)\{1 - F_0(t - \beta'_0x)\}} 1_{[\epsilon, 1-\epsilon]}(F_0(t - \beta'_0x)).$$

So we get, using [1],

$$\sup_{\zeta > 0} \zeta H(\zeta, \mathcal{K}'_2, \mathbb{P}_n) = O_p(1),$$

implying

$$\int_0^\zeta H(u, \mathcal{K}'_2, \mathbb{P}_n)^{1/2} du = O_p(\zeta^{1/2}), \quad \zeta > 0.$$

As before, we now get:

$$\begin{aligned} & \int_{\hat{F}_{n, \hat{\beta}_n}(t - \hat{\beta}'_n x) \in [\epsilon, 1-\epsilon]} \left\{ x f_{nh, \hat{\beta}_n}(t - \hat{\beta}'_n x) - \varphi_{\hat{\beta}_n}(t - \hat{\beta}'_n x) \right\} \\ & \quad \cdot \frac{\hat{F}_{n, \hat{\beta}_n}(t - \hat{\beta}'_n x) - \delta}{\hat{F}_{n, \hat{\beta}_n}(t - \hat{\beta}'_n x)\{1 - \hat{F}_{n, \hat{\beta}_n}(t - \hat{\beta}'_n x)\}} d(\mathbb{P}_n - P_0)(t, x, \delta) \\ &= \int_{F_0(t - \beta'_0 x) \in [\epsilon, 1-\epsilon]} \left\{ x f_0(t - \beta'_0 x) - \varphi_{\beta_0}(t - \beta'_0 x) \right\} \\ & \quad \cdot \frac{F_0(t - \beta'_0 x) - \delta}{F_0(t - \beta'_0 x)\{1 - F_0(t - \beta'_0 x)\}} d(\mathbb{P}_n - P_0)(t, x, \delta) \\ & \quad + o_p\left(n^{-1/2} + \hat{\beta}_n - \beta_0\right) \end{aligned}$$

and

$$\begin{aligned}
& \int_{\hat{F}_{n,\hat{\beta}_n}(t-\hat{\beta}'_n x) \in [\epsilon, 1-\epsilon]} \left\{ x f_{nh,\hat{\beta}_n}(t - \hat{\beta}'_n x) - \varphi_{\hat{\beta}_n}(t - \hat{\beta}'_n x) \right\} \\
& \quad \cdot \frac{\hat{F}_{n,\hat{\beta}_n}(t - \hat{\beta}'_n x) - F_0(t - \beta'_0 x)}{\hat{F}_{n,\hat{\beta}_n}(t - \hat{\beta}'_n x) \{1 - \hat{F}_{n,\hat{\beta}_n}(t - \hat{\beta}'_n x)\}} dP_0(t, x, \delta) \\
& = \left\{ \int_{F_0(t-\beta'_0 x) \in [\epsilon, 1-\epsilon]} \left\{ x f_0(t - \beta'_0 x) - \varphi_{\beta_0}(t - \beta'_0 x) \right\} \right. \\
& \quad \cdot \frac{f_0(t - \beta'_0 x) x'}{F_0(t - \beta'_0 x) \{1 - F_0(t - \beta'_0 x)\}} dP_0(t, x, \delta) \left. \right\} (\hat{\beta}_n - \beta_0) \\
& \quad + o_p \left(n^{-1/2} + \hat{\beta}_n - \beta_0 \right).
\end{aligned}$$

The result now follows. \square

S12. Asymptotic behavior of the plug-in estimator. In this section we first sketch in Section S12.1 the proof of consistency of the plug-in estimator, denoted by $\hat{\beta}_n$. This is the second result stated in Theorem 4.3. The proof of existence of a root is similar to the proof of existence of a root of the simple score estimator defined in Section 4.1 and omitted. We next prove the asymptotic normality result of the plug-in estimator, which is the third result given in Theorem 4.3. The proof of Theorem 4.5 on the asymptotic representation of the plug-in estimator as a sum of i.i.d. random variables follows from the proof of 4.3. The asymptotic distribution of the estimator of the intercept, given in Theorem 5.1, is proved in Section S12.2.

S12.1. Consistency and asymptotic normality of the plug-in estimator. We start by proving that $\hat{\beta}_n$ is a consistent estimate of β_0 .

PROOF OF THEOREM 4.3, PART 1 (CONSISTENCY OF THE PLUG-IN ESTIMATOR). We assume that $\hat{\beta}_n$ is contained in the compact set Θ , and hence the sequence $(\hat{\beta}_n)$ has a subsequence $(\hat{\beta}_{n_k} = \hat{\beta}_{n_k}(\omega))$, converging to an element β_* . It is easily seen that, if $\hat{\beta}_{n_k} = \hat{\beta}_{n_k}(\omega) \rightarrow \beta_*$, we get:

$$F_{n_k, \hat{\beta}_{n_k}}(t - \hat{\beta}'_{n_k} x) \rightarrow F_{\beta_*}(t - \beta'_* x) \stackrel{\text{def}}{=} \int F_0(t - \beta'_* x + (\beta_* - \beta_0)' y) f_{X|T-\beta'_* X}(y|t - \beta'_* x) dy.$$

In the limit we get therefore the relation

$$\begin{aligned}
& \lim_{k \rightarrow \infty} -(\hat{\beta}_{n_k} - \beta_0)' \int_{F_{n_k, \hat{\beta}_{n_k}}(t - \hat{\beta}'_{n_k} x) \in [\epsilon, 1 - \epsilon]} \frac{\{\delta - F_{n_k, \hat{\beta}_{n_k}}(t - \hat{\beta}'_{n_k} x)\} \partial_\beta F_{n_k, \beta}(t - \beta' x)|_{\beta = \hat{\beta}_{n_k}}}{F_{n_k, \hat{\beta}_{n_k}}(t - \hat{\beta}'_{n_k} x) \{1 - F_{n_k, \hat{\beta}_{n_k}}(t - \hat{\beta}'_{n_k} x)\}} d\mathbb{P}_{n_k}(t, x, \delta) \\
& \text{(S12.1)} \\
& = -(\beta_* - \beta_0)' \int_{F_{\beta_*}(t - \beta'_* x) \in [\epsilon, 1 - \epsilon]} \frac{\{F_0(t - \beta'_0 x) - F_{\beta_*}(t - \beta'_* x)\} \partial_\beta F_\beta(t - \beta' x)|_{\beta = \beta_*}}{F_{\beta_*}(t - \beta'_* x) \{1 - F_{\beta_*}(t - \beta'_* x)\}} dG(t, x) = 0,
\end{aligned}$$

which can only mean $\beta_* = \beta_0$ by condition (4.12). \square

We continue with the asymptotic normality proof for the plug-in estimator. Before we start the proof, we give some auxiliary results on the L_2 -distance between the plug-in estimate $F_{nh, \beta}$ and F_β and between the partial derivative of the plug-in estimate $\partial_\beta F_{nh, \beta}$ and $\partial_\beta F_\beta$ in Lemma S12.1. For simplicity, we derive the proof of Lemma S12.1 for the one-dimensional case and let $\Theta = [\beta_0 - \eta, \beta_0 + \eta]$ for some $\eta > 0$. The higher-dimensional extension of the one-dimensional proof is straightforward. Next, we follow the arguments used to prove the asymptotic normality of the estimators defined in Theorem 4.1 and Theorem 4.2 and give a similar proof for the limiting distribution of the plug-in estimator.

LEMMA S12.1. *Let the conditions of Theorem 4.3 be satisfied and let $k = 1$. Let the function F_β be defined by (3.2). Then we have, for the estimate $F_{nh, \beta}$, defined by (4.7),*

$$\text{(S12.2)} \quad \int_{F_{nh, \beta}(t - \beta x) \in [\epsilon, 1 - \epsilon]} \{F_{nh, \beta}(t - \beta x) - F_\beta(t - \beta x)\}^2 dG(t, x) = O_p\left(\frac{1}{nh}\right) + O_p(h^4),$$

$$\text{(S12.3)} \quad \int_{F_{nh, \beta}(t - \beta x) \in [\epsilon, 1 - \epsilon]} \{\partial_\beta F_{nh, \beta}(t - \beta x) - \partial_\beta F_\beta(t - \beta x)\}^2 dG(t, x) = O_p\left(\frac{1}{nh^3}\right) + O_p(h^2)$$

uniformly in $\beta \in [\beta_0 - \eta, \beta_0 + \eta]$. The results remain valid when dG in (S12.2) or (S12.3) is replaced by $d\mathbb{G}_n$.

PROOF OF LEMMA S12.1. We first prove the first part and show that (S12.2) holds. Recall that,

$$F_{nh, \beta}(t - \beta x) = \frac{g_{nh, 1, \beta}(t - \beta x)}{g_{nh, \beta}(t - \beta x)}$$

where

$$g_{nh,1,\beta}(t - \beta x) = \int \delta K_h(t - \beta x - u + \beta y) d\mathbb{P}_n(u, y, \delta),$$

and

$$g_{nh,\beta}(t - \beta x) = \int K_h(t - \beta x - u + \beta y) d\mathbb{P}_n(u, y, \delta).$$

Moreover,

$$F_\beta(t - \beta x) = \int F_0(t - \beta_0 x + (\beta - \beta_0)(y - x)) f_{X|T-\beta X}(y|t - \beta x) dy.$$

We first investigate the bias part.

$$\begin{aligned} \mathbb{E}g_{nh,1,\beta}(t - \beta x) &= \int F_0(u - \beta_0 y) K_h(t - \beta x - u + \beta y) dG(u, y) \\ &= \int F_0(v + (\beta - \beta_0)y) K_h(t - \beta x - v) f_{T-\beta X}(v) f_{X|T-\beta X}(y|v) dy dv \\ &= \int F_0(t - \beta x + (\beta - \beta_0)y - hw) K(w) f_{T-\beta X}(t - \beta x - hw) f_{X|T-\beta X}(y|t - \beta x - hw) dy dw \\ &= f_{T-\beta X}(t - \beta x) \int F_0(t - \beta_0 x + (\beta - \beta_0)(y - x)) f_{X|T-\beta X}(y|t - \beta x) dy + O(h^2), \end{aligned}$$

uniformly in $\beta \in [\beta_0 - \eta, \beta_0 + \eta]$ and t, x varying over a finite interval, due to the assumptions of Theorem 4.3. In a similar way, we get

$$\mathbb{E}g_{nh,\beta}(t - \beta x) = f_{T-\beta X}(t - \beta x) + O(h^2),$$

uniformly in $\beta \in [\beta_0 - \eta, \beta_0 + \eta]$ and t, x varying over a finite interval. So we find:

$$\frac{\mathbb{E}g_{nh,1,\beta}(t - \beta x)}{\mathbb{E}g_{nh,\beta}(t - \beta x)} = F_\beta(t - \beta x) + O(h^2).$$

uniformly in $\beta \in [\beta_0 - \eta, \beta_0 + \eta]$ and t, x varying over a finite interval, such that $\mathbb{E}g_{nh,1,\beta}(t - \beta x)$ stays away from zero.

So we obtain

$$\begin{aligned} &F_{nh,\beta}(t - \beta x) - F_\beta(t - \beta x) \\ &= \frac{g_{nh,1,\beta}(t - \beta x) - \mathbb{E}g_{nh,1,\beta}(t - \beta x)}{g_{nh,\beta}(t - \beta x)} + \mathbb{E}g_{nh,1,\beta}(t - \beta x) \frac{\mathbb{E}g_{nh,\beta}(t - \beta x) - g_{nh,\beta}(t - \beta x)}{g_{nh,\beta}(t - \beta x)\mathbb{E}g_{nh,\beta}(t - \beta x)} + O(h^2), \end{aligned}$$

and

$$\begin{aligned}
& \{F_{nh,\beta}(t - \beta x) - F_\beta(t - \beta x)\}^2 \\
& \leq 3 \left\{ \frac{g_{nh,1,\beta}(t - \beta x) - \mathbb{E}g_{nh,1,\beta}(t - \beta x)}{g_{nh,\beta}(t - \beta x)} \right\}^2 + 3 \left\{ \mathbb{E}g_{nh,1,\beta}(t - \beta x) \frac{\mathbb{E}g_{nh,\beta}(t - \beta x) - g_{nh,\beta}(t - \beta x)}{g_{nh,\beta}(t - \beta x)\mathbb{E}g_{nh,\beta}(t - \beta x)} \right\}^2 \\
& \quad (S12.4) \\
& \quad \quad \quad + O(h^4).
\end{aligned}$$

uniformly in $\beta \in [\beta_0 - \eta, \beta_0 + \eta]$ and t, x varying over a finite interval, such that $\mathbb{E}g_{nh,1,\beta}(t - \beta x)$ stays away from zero.

Since $\eta > 0$ is chosen in such a way that $a_1(\beta) = F_\beta^{-1}(\epsilon) > a$, $b_1(\beta) = F_\beta^{-1}(1 - \epsilon) < b$, for each $\beta \in [\beta_0 - \eta, \beta_0 + \eta]$ and since $g_{nh,\beta}$ stays away from zero with probability tending to one if $\epsilon < F_{nh,\beta}(t - \beta x) < 1 - \epsilon$ we get

$$\begin{aligned}
& \int_{F_{nh,\beta}(t-\beta x) \in [\epsilon, 1-\epsilon]} \left\{ \frac{g_{nh,1,\beta}(t - \beta x) - \mathbb{E}g_{nh,1,\beta}(t - \beta x)}{g_{nh,\beta}(t - \beta x)} \right\}^2 dG(t, x) \\
& \lesssim \int_{F_{nh,\beta}(t-\beta x) \in [\epsilon, 1-\epsilon]} \{g_{nh,1,\beta}(t - \beta x) - \mathbb{E}g_{nh,1,\beta}(t - \beta x)\}^2 dG(t, x)
\end{aligned}$$

Furthermore

$$\begin{aligned}
\mathbb{E} \{g_{nh,1,\beta}(t - \beta x) - \mathbb{E}g_{nh,1,\beta}(t - \beta x)\}^2 &= \mathbb{E} \left\{ \int \delta K_h(t - \beta x - u + \beta y) d(\mathbb{P}_n - P_0)(u, y, \delta) \right\}^2 \\
&= O\left(\frac{1}{nh}\right),
\end{aligned}$$

uniformly for (t, x) in a bounded region, so we get

$$\mathbb{E} \int_{F_{nh,\beta}(t-\beta x) \in [\epsilon, 1-\epsilon]} \{g_{nh,1,\beta}(t - \beta x) - \mathbb{E}g_{nh,1,\beta}(t - \beta x)\}^2 dG(t, x) = O\left(\frac{1}{nh}\right).$$

Hence

$$\int_{F_{nh,\beta}(t-\beta x) \in [\epsilon, 1-\epsilon]} \left\{ \frac{g_{nh,1,\beta}(t - \beta x) - \mathbb{E}g_{nh,1,\beta}(t - \beta x)}{g_{nh,\beta}(t - \beta x)} \right\}^2 dG(t, x) = O_p\left(\frac{1}{nh}\right).$$

The second term on the right-hand side of (S12.4) can be treated in a similar way. So we get (S12.2). This proves (S12.2).

We next replace dG in part (S12.2) by $d\mathbb{G}_n$ and we get

$$\begin{aligned} & \int_{F_{nh,\beta}(t-\beta x) \in [\epsilon, 1-\epsilon]} \left\{ \frac{g_{nh,1,\beta}(t-\beta x) - \mathbb{E}g_{nh,1,\beta}(t-\beta x)}{g_{nh,\beta}(t-\beta x)} \right\}^2 d\mathbb{G}_n(t, x) \\ & \lesssim \int_{F_{nh,\beta}(t-\beta x) \in [\epsilon, 1-\epsilon]} \{g_{nh,1,\beta}(t-\beta x) - \mathbb{E}g_{nh,1,\beta}(t-\beta x)\}^2 d\mathbb{G}_n(t, x) \\ & = \frac{1}{n} \sum_{i=1}^n \{g_{nh,1,\beta}(T_i - \beta X_i) - \mathbb{E}g_{nh,1,\beta}(T_i - \beta X_i)\}^2 1_{\{\epsilon < F_{nh,\beta}(T_i - \beta X_i) < 1-\epsilon\}}. \end{aligned}$$

Moreover,

$$\begin{aligned} & \mathbb{E} \frac{1}{n} \sum_{i=1}^n \{g_{nh,1,\beta}(T_i - \beta X_i) - \mathbb{E}g_{nh,1,\beta}(T_i - \beta X_i)\}^2 1_{\{\epsilon < F_{nh,\beta}(T_i - \beta X_i) < 1-\epsilon\}} \\ & = \mathbb{E} \{g_{nh,1,\beta}(T_1 - \beta X_1) - \mathbb{E}g_{nh,1,\beta}(T_1 - \beta X_1)\}^2 1_{\{\epsilon < F_{nh,\beta}(T_1 - \beta X_1) < 1-\epsilon\}} \\ & \lesssim \mathbb{E} \int_{\epsilon/2 < F_\beta(t-\beta x) < 1-\epsilon/2} \{g_{nh,1,\beta}(t-\beta x) - \mathbb{E}g_{nh,1,\beta}(t-\beta x)\}^2 dG(t, x) \\ & = O\left(\frac{1}{nh}\right). \end{aligned}$$

This implies by the Markov inequality,

$$\int_{F_{nh,\beta}(t-\beta x) \in [\epsilon, 1-\epsilon]} \left\{ \frac{g_{nh,1,\beta}(t-\beta x) - \mathbb{E}g_{nh,1,\beta}(t-\beta x)}{g_{nh,\beta}(t-\beta x)} \right\}^2 d\mathbb{G}_n(t, x) = O_p\left(\frac{1}{nh}\right).$$

The other term on the right-hand side of (S12.4) is treated similarly; and the result of (S12.2) also follows when we replace dG by $d\mathbb{G}_n$.

We next continue with the proof of (S12.3).

We have:

$$(S12.5) \quad \partial_\beta F_{nh,\beta}(t-\beta x) = \frac{\int (y-x) \{\delta - F_{nh,\beta}(t-\beta x)\} K'_h(t-\beta x-u+\beta y) d\mathbb{P}_n(u, y, \delta)}{g_{nh,\beta}(t-\beta x)}.$$

We consider the numerator of (S12.5). It can be rewritten as

$$\begin{aligned} & \int (y-x) \{\delta - F_0(u-\beta_0 y)\} K'_h(t-\beta x-u+\beta y) d\mathbb{P}_n(u, y, \delta) \\ & + \int (y-x) \{F_0(u-\beta_0 y) - F_\beta(t-\beta x)\} K'_h(t-\beta x-u+\beta y) d\mathbb{G}_n(u, y) \\ & + \{F_\beta(t-\beta x) - F_{nh,\beta}(t-\beta x)\} \int (y-x) K'_h(t-\beta x-u+\beta y) d\mathbb{G}_n(u, y). \end{aligned}$$

The first term can be written as

$$A_n(t, x, \beta) \stackrel{\text{def}}{=} \int (y - x) \{ \delta - F_0(u - \beta_0 y) \} K'_h(t - \beta x - u + \beta y) d(\mathbb{P}_n - P_0)(u, y, \delta),$$

and we have:

$$\begin{aligned} \mathbb{E} \int_{F_{nh, \beta}(t - \beta x) \in [\epsilon, 1 - \epsilon]} A_n(t, x, \beta)^2 dG(t, x) &\leq \mathbb{E} \int A_n(t, x, \beta)^2 dG(t, x) \\ &\sim \frac{1}{nh^3} \int \text{var}(X|v) F_0(v) \{1 - F_0(v)\} f_{T - \beta X}(v) dv \int K'(u)^2 du, \quad n \rightarrow \infty. \end{aligned}$$

In the second term we must compare $F_0(u - \beta_0 y)$ with

$$F_\beta(t - \beta x) = \int F_0(t - \beta_0 x + (\beta - \beta_0)(z - x)) f_{X|T - \beta X}(z|t - \beta x) dz.$$

We can write

$$\begin{aligned} &F_0(u - \beta_0 y) - F_\beta(t - \beta x) \\ &= \int \{F_0(u - \beta_0 y) - F_0(t - \beta_0 x + (\beta - \beta_0)(z - x))\} f_{X|T - \beta X}(z|t - \beta x) dz. \end{aligned}$$

So we find for the second term

$$\begin{aligned} B_n(t, x, \beta) &\stackrel{\text{def}}{=} \int (y - x) \{F_0(u - \beta_0 y) - F_\beta(t - \beta x)\} K'_h(t - \beta x - u + \beta y) d\mathbb{G}_n(u, y) \\ &= \int \int (y - x) \{F_0(u - \beta_0 y) - F_0(t - \beta_0 x + (\beta - \beta_0)(z - x))\} f_{X|T - \beta X}(z|t - \beta x) dz \\ &\quad \cdot K'_h(t - \beta x - u + \beta y) d\mathbb{G}_n(u, y) \\ &= \int (y - x) \int \{F_0(u - \beta_0 y) - F_0(t - \beta_0 x + (\beta - \beta_0)(z - x))\} f_{X|T - \beta X}(z|t - \beta x) dz \\ &\quad \cdot K'_h(t - \beta x - u + \beta y) dG(u, y) \\ &\quad + \int (y - x) \int \{F_0(u - \beta_0 y) - F_0(t - \beta_0 x + (\beta - \beta_0)(z - x))\} f_{X|T - \beta X}(z|t - \beta x) dz \\ &\quad \cdot K'_h(t - \beta x - u + \beta y) d(\mathbb{G}_n - G)(u, y) \\ &= f_{T - \beta X}(t - \beta x) \partial_\beta F_\beta(t - \beta x) + O(h) + O_p\left(\frac{1}{nh^3}\right). \end{aligned}$$

where, using integration by parts, the last line follows by straightforward calculation.

Since

$$g_{nh, \beta}(t - \beta x) = f_{T - \beta X}(t - \beta x) + O_p(h^2),$$

we get,

$$\int_{F_{nh,\beta}(t-\beta x) \in [\epsilon, 1-\epsilon]} \left\{ \frac{B_n(t, x, \beta)}{g_{nh,\beta}(t-\beta x)} - \partial_\beta F_\beta(t-\beta x) \right\}^2 dG(t, x) = O_p\left(\frac{1}{nh^3}\right) + O_p(h^2).$$

Finally, defining

$$C_n(t, x, \beta) \stackrel{\text{def}}{=} \{F_\beta(t-\beta x) - F_{nh,\beta}(t-\beta x)\} \int (y-x) K'_h(t-\beta x-u+\beta y) d\mathbb{G}_n(u, y),$$

we get, using,

$$\begin{aligned} & \int (y-x) K'_h(t-\beta x-u+\beta y) d\mathbb{G}_n(u, y) \\ &= \int (y-x) K'_h(t-\beta x-u+\beta y) dG(u, y) + \int (y-x) K'_h(t-\beta x-u+\beta y) d(\mathbb{G}_n - G)(u, y) \\ &= \int (y-x) K'_h(t-\beta x-v) f_{T-\beta X}(v) f_{X|T-\beta X}(y|v) dv dy + O_p\left(\frac{1}{nh^3}\right) \\ &= \int (y-x) K_h(t-\beta x-v) \frac{d}{dv} \{f_{T-\beta X}(v) f_{X|T-\beta X}(y|v)\} dv dy + O_p\left(\frac{1}{nh^3}\right) \\ &= O_p(1), \end{aligned}$$

and using the first part of Lemma [S12.1](#) for the factor $F_\beta(t-\beta x) - F_{nh,\beta}(t-\beta x)$ that

$$\int_{F_{nh,\beta}(t-\beta x) \in [\epsilon, 1-\epsilon]} C_n(t, x, \beta)^2 dG(t, x) = O_p\left(\frac{1}{nh}\right) + O_p(h^4).$$

This proves [\(S12.3\)](#). The second part of the result, replacing dG by $d\mathbb{G}_n$ in [\(S12.3\)](#) is proved in the same way as the second part of [\(S12.2\)](#). \square

PROOF OF THEOREM [4.3](#), PART 2 (ASYMPTOTIC NORMALITY OF THE PLUG-IN ESTIMATOR).

To prove the asymptotic normality of the plug-in estimator, we follow the reasoning of the corresponding proofs of the simple score estimator and the efficient score estimator described in Section [4.1](#) and Section [4.2](#). We prove that,

(S12.6)

$$\begin{aligned} & \psi_{3,nh}^{(\epsilon)}(\hat{\beta}_n) \\ &= \int_{F_0(t-\beta'_0 x) \in [\epsilon, 1-\epsilon]} \frac{\{E(X|T-\beta'_0 X = t-\beta'_0 x) - x\} f_0(t-\beta'_0 x) \{F_0(t-\beta'_0 x) - \delta\}}{F_0(t-\beta'_0 x) \{1 - F_0(t-\beta'_0 x)\}} d\mathbb{P}_n(t, x, \delta) \\ & \quad + \psi'_{3,\epsilon}(\beta_0)(\hat{\beta}_n - \beta_0) + o_p\left(n^{-1/2} + (\hat{\beta}_n - \beta_0)\right), \end{aligned}$$

where $\psi_{3,\epsilon}$ is defined by,

$$(S12.7) \quad \psi_{3,\epsilon}(\beta) = \int_{F_\beta(t-\beta'x) \in [\epsilon, 1-\epsilon]} \frac{\partial_\beta F_\beta(t-\beta'x) \{F_\beta(t-\beta'x) - \delta\}}{F_\beta(t-\beta'x) \{1 - F_\beta(t-\beta'x)\}} dP_0(t, x, \delta),$$

and

$$\psi'_{3,\epsilon}(\beta_0) = \mathbb{E}_\epsilon \left\{ \frac{f_0(T - \beta'_0 X)^2 \{X - \mathbb{E}(X|T - \beta'_0 X)\} \{X - \mathbb{E}(X|T - \beta'_0 X)\}'}{F_0(T - \beta'_0 X) \{1 - F_0(T - \beta'_0 X)\}} \right\} = I_\epsilon(\beta_0),$$

which follows by straightforward calculations after noting that,

$$\begin{aligned} \partial_\beta F_\beta(t - \beta'x) &= \int (y - x) f_0(t - \beta'_0 x + (\beta - \beta_0)'(y - x)) f_{X|T-\beta'X}(y|T - \beta'X = t - \beta'x) dy \\ &\quad + \int F_0(t - \beta'_0 x + (\beta - \beta_0)'(y - x)) \partial_\beta f_{X|T-\beta'X}(y|T - \beta'X = t - \beta'x) dG(t, x) \end{aligned}$$

is, at $\beta = \beta_0$ equal to

$$f_0(t - \beta'_0 x) \mathbb{E} \{X - x | T - \beta'_0 X = t - \beta'_0 x\}.$$

We have,

$$\begin{aligned} \psi_{3,nh}^{(\epsilon)}(\hat{\beta}_n) &= \int_{F_{nh,\hat{\beta}_n}(t-\hat{\beta}'_n x) \in [\epsilon, 1-\epsilon]} \partial_\beta F_{nh,\beta}(t - \beta'x) \big|_{\beta=\hat{\beta}_n} \frac{F_{nh,\hat{\beta}_n}(t - \hat{\beta}'_n x) - \delta}{F_{nh,\hat{\beta}_n}(t - \hat{\beta}'_n x) \{1 - F_{nh,\hat{\beta}_n}(t - \hat{\beta}'_n x)\}} d\mathbb{P}_n(t, x, \delta) \\ &= \int_{F_{nh,\hat{\beta}_n}(t-\hat{\beta}'_n x) \in [\epsilon, 1-\epsilon]} \partial_\beta F_\beta(t - \beta'x) \big|_{\beta=\hat{\beta}_n} \frac{F_{nh,\hat{\beta}_n}(t - \hat{\beta}'_n x) - \delta}{F_{nh,\hat{\beta}_n}(t - \hat{\beta}'_n x) \{1 - F_{nh,\hat{\beta}_n}(t - \hat{\beta}'_n x)\}} d\mathbb{P}_n(t, x, \delta) \\ &\quad + \int_{F_{nh,\hat{\beta}_n}(t-\hat{\beta}'_n x) \in [\epsilon, 1-\epsilon]} \left\{ \partial_\beta F_{nh,\beta}(t - \beta'x) \big|_{\beta=\hat{\beta}_n} - \partial_\beta F_\beta(t - \beta'x) \big|_{\beta=\hat{\beta}_n} \right\} \\ &\quad \cdot \frac{F_{nh,\hat{\beta}_n}(t - \hat{\beta}'_n x) - \delta}{F_{nh,\hat{\beta}_n}(t - \hat{\beta}'_n x) \{1 - F_{nh,\hat{\beta}_n}(t - \hat{\beta}'_n x)\}} d\mathbb{P}_n(t, x, \delta) \\ &= I + II \end{aligned}$$

Let \mathcal{F} be a class of functions with the property that

$$\int_{\epsilon/2 < F_\beta(u) < 1-\epsilon/2} f'(u)^2 du \leq M.$$

if $f \in \mathcal{F}$, for a fixed $M > 0$. Using Proposition 5.1.9, p. 393 in [11], with $m = 1$, $p = 2$ and $h \asymp n^{-1/5}$, we may assume that the functions $u \rightarrow F_{nh,\beta}(u)$ and $u \rightarrow$

$\partial_\beta F_{nh,\beta}(u)$ belong to \mathcal{F} . Since the plug-in estimates are monotonically increasing with probability tending to one we get that the function

$$(t, x) \mapsto 1_{[\epsilon, 1-\epsilon]}(F_{nh,\beta}(t - \beta'x)),$$

can be written in the form

$$(t, x) \mapsto 1_{[a_{\epsilon, F_{nh,\beta}}, b_{\epsilon, F_{nh,\beta}}]}(t - \beta'x) = 1_{[a_{\epsilon, F_{nh,\beta}}, \infty)}(t - \beta'x) - 1_{(b_{\epsilon, F_{nh,\beta}}, \infty)}(t - \beta'x),$$

for $a_{\epsilon, F_{nh,\beta}} \leq b_{\epsilon, F_{nh,\beta}}$ for large n , with probability tending to one. The function is therefore of uniformly bounded variation for n sufficiently large (see also the proofs of Theorems 4.1 and 4.2). It now follows from [1] that the random ζ -entropy $H(\zeta, \mathcal{K}_3, \mathbb{P}_n)$ for the class \mathcal{K}_3 of functions consisting of the function which is identically zero and the functions

$$\begin{aligned} \mathcal{K}_3 = \left\{ (t, x, \delta) \mapsto \left\{ \partial_\beta F_{nh,\beta}(t - \beta'x) - \partial_\beta F_\beta(t - \beta'x) \right\} \frac{F(t - \beta'x) - \delta}{F(t - \beta'x)\{1 - F(t - \beta'x)\}} \right. \\ \left. \cdot 1_{[\epsilon, 1-\epsilon]}(F_{nh,\beta}(t - \beta'x)) : F \in \mathcal{F}, \beta \in \Theta \right\}, \end{aligned} \quad (\text{S12.8})$$

w.r.t. the L_2 -distance d_n , defined by (S11.7) satisfies:

$$\sup_{\zeta > 0} \zeta H(\zeta, \mathcal{K}_3, \mathbb{P}_n) = O_p(1),$$

implying:

$$\int_0^\zeta H(u, \mathcal{K}_3, \mathbb{P}_n)^{1/2} du = O_p(\zeta^{1/2}), \quad \zeta > 0.$$

Moreover, by Lemma S12.1 we also have,

$$\begin{aligned} \int_{F_{nh,\beta}(t - \beta'x) \in [\epsilon, 1-\epsilon]} \left\{ \left\{ \partial_\beta F_{nh,\beta}(t - \beta'x) - \partial_\beta F_\beta(t - \beta'x) \right\} \right. \\ \left. \cdot \frac{F_{nh,\beta}(t - \beta'x) - \delta}{F_{nh,\beta}(t - \beta'x)\{1 - F_{nh,\beta}(t - \beta'x)\}} \right\}^2 dP_0(t, x, \delta) \xrightarrow{p} 0. \end{aligned}$$

This implies by an application of Lemma 8.1, that,

$$\begin{aligned} \int_{F_{nh,\hat{\beta}_n}(t - \hat{\beta}'_n x) \in [\epsilon, 1-\epsilon]} \left\{ \left\{ \partial_\beta F_{nh,\beta}(t - \beta'x) \big|_{\beta=\hat{\beta}_n} - \partial_\beta F_\beta(t - \beta'x) \big|_{\beta=\hat{\beta}_n} \right\} \right. \\ \left. \cdot \frac{F_{nh,\hat{\beta}_n}(t - \hat{\beta}'_n x) - \delta}{F_{nh,\hat{\beta}_n}(t - \hat{\beta}'_n x)\{1 - F_{nh,\hat{\beta}_n}(t - \hat{\beta}'_n x)\}} \right\} d(\mathbb{P}_n - P_0)(t, x, \delta) = o_p(n^{-1/2}) \end{aligned}$$

Furthermore, an application of the Cauchy-Schwarz inequality and Lemma [S12.1](#) yield that

$$\begin{aligned} & \sqrt{n} \int_{F_{nh, \hat{\beta}_n}(t - \beta x) \in [\epsilon, 1 - \epsilon]} \left\{ \partial_\beta F_\beta(t - \beta' x) \big|_{\beta = \hat{\beta}_n} - \partial_\beta F_{nh, \beta}(t - \beta' x) \big|_{\beta = \hat{\beta}_n} \right\} \\ & \quad \cdot \left\{ \frac{F_{nh, \hat{\beta}_n}(t - \hat{\beta}'_n x) - F_0(t - \beta'_0 x)}{F_{nh, \hat{\beta}_n}(t - \hat{\beta}'_n x) \{1 - F_{nh, \hat{\beta}_n}(t - \hat{\beta}'_n x)\}} \right\} dP_0(t, x, \delta) \\ & = O_p\left(n^{-1/10}\right) + o_p\left(\sqrt{n}(\hat{\beta}_n - \beta_0)\right) \end{aligned}$$

The conclusion is that

$$II = o_p\left(n^{-1/2} + (\hat{\beta}_n - \beta_0)\right)$$

We now write:

$$\begin{aligned} I &= \int_{F_{nh, \hat{\beta}_n}(t - \hat{\beta}'_n x) \in [\epsilon, 1 - \epsilon]} \partial_\beta F_\beta(t - \beta' x) \big|_{\beta = \hat{\beta}_n} \\ & \quad \cdot \frac{F_{nh, \hat{\beta}_n}(t - \hat{\beta}'_n x) - \delta}{F_{nh, \hat{\beta}_n}(t - \hat{\beta}'_n x) \{1 - F_{nh, \hat{\beta}_n}(t - \hat{\beta}'_n x)\}} d\mathbb{P}_n(t, x, \delta) \\ &= \int_{F_{nh, \hat{\beta}_n}(t - \hat{\beta}'_n x) \in [\epsilon, 1 - \epsilon]} \partial_\beta F_\beta(t - \beta' x) \big|_{\beta = \hat{\beta}_n} \\ & \quad \cdot \frac{F_{\hat{\beta}_n}(t - \hat{\beta}'_n x) - \delta}{F_{n, \hat{\beta}_n}(t - \hat{\beta}'_n x) \{1 - F_{n, \hat{\beta}_n}(t - \hat{\beta}'_n x)\}} d\mathbb{P}_n(t, x, \delta) \\ & \quad + \int_{F_{nh, \hat{\beta}_n}(t - \hat{\beta}'_n x) \in [\epsilon, 1 - \epsilon]} \partial_\beta F_\beta(t - \beta' x) \big|_{\beta = \hat{\beta}_n} \\ & \quad \cdot \frac{F_{n, \hat{\beta}_n}(t - \hat{\beta}'_n x) - F_{\hat{\beta}_n}(t - \hat{\beta}'_n x)}{\hat{F}_{\hat{\beta}_n}(t - \hat{\beta}'_n x) \{1 - F_{n, \hat{\beta}_n}(t - \hat{\beta}'_n x)\}} d\mathbb{P}_n(t, x, \delta) \\ &= I_a + I_b. \end{aligned}$$

We now get, using Lemma [S12.1](#) and

$$\partial_\beta F_\beta(t - \beta' x) \big|_{\beta = \hat{\beta}_n} = E(X|T - \beta'_0 X = t - \beta'_0 x) f_0(t - \beta'_0 x) + O_p\left(\hat{\beta}_n - \beta_0\right),$$

that $I_b = o_p\left(n^{-1/2} + \hat{\beta}_n - \beta_0\right)$. The result of Theorem [4.3](#) now follows by showing

that,

$$\begin{aligned}
 & \int_{F_{nh, \hat{\beta}_n}(t - \hat{\beta}'_n x) \in [\epsilon, 1 - \epsilon]} \partial_\beta F_\beta(t - \beta' x) \big|_{\beta = \hat{\beta}_n} \\
 & \quad \cdot \frac{F_{\hat{\beta}_n}(t - \hat{\beta}'_n x) - \delta}{F_{nh, \hat{\beta}_n}(t - \hat{\beta}'_n x) \{1 - F_{nh, \hat{\beta}_n}(t - \hat{\beta}'_n x)\}} d(\mathbb{P}_n - P_0)(t, x, \delta) \\
 & = \int_{F_0(t - \beta'_0 x) \in [\epsilon, 1 - \epsilon]} \partial_\beta F_\beta(t - \beta' x) \big|_{\beta = \beta_0} \\
 & \quad \cdot \frac{F_0(t - \beta'_0 x) - \delta}{F_0(t - \beta'_0 x) \{1 - F_0(t - \beta'_0 x)\}} d(\mathbb{P}_n - P_0)(t, x, \delta) \\
 & \quad + o_p \left(n^{-1/2} + \hat{\beta}_n - \beta_0 \right)
 \end{aligned}
 \tag{S12.9}$$

and,

$$\begin{aligned}
 & \int_{F_{nh, \hat{\beta}_n}(t - \hat{\beta}'_n x) \in [\epsilon, 1 - \epsilon]} \partial_\beta F_\beta(t - \beta' x) \big|_{\beta = \hat{\beta}_n} \\
 & \quad \cdot \frac{F_{\hat{\beta}_n}(t - \hat{\beta}'_n x) - \delta}{F_{nh, \hat{\beta}_n}(t - \hat{\beta}'_n x) \{1 - F_{nh, \hat{\beta}_n}(t - \hat{\beta}'_n x)\}} dP_0(t, x, \delta) \\
 & = \psi'_{3, \epsilon}(\beta_0)(\hat{\beta}_n - \beta_0) + o_p \left(n^{-1/2} + \hat{\beta}_n - \beta_0 \right).
 \end{aligned}
 \tag{S12.10}$$

The proof of (S12.9) and (S12.10) is similar to the proof of the corresponding steps given in the proof of Theorem 4.1 and omitted. \square

REMARK S12.1. It follows from the proof of Theorem 4.3 that

$$\begin{aligned}
 & \sqrt{n} I_\epsilon(\beta_0)(\hat{\beta}_n - \beta_0) \\
 & = n^{-1/2} \sum_{i=1}^n f_0(T_i - \beta'_0 X_i) \{E_{\beta_0}(X_i | T_i - \beta'_0 X_i) - X_i\} \\
 & \quad \cdot \frac{\Delta_i - F_0(T_i - \beta'_0 X_i)}{F_0(T_i - \beta'_0 X_i) \{1 - F_0(T_i - \beta'_0 X_i)\}} 1_{[\epsilon, 1 - \epsilon]} \{F_0(T_i - \beta'_0 X_i)\} + o_p(1).
 \end{aligned}$$

Therefore the result of Theorem 4.5 follows.

S12.2. *Estimation of the intercept.*

PROOF OF THEOREM 5.1. We will denote $dx_1 \dots dx_k$ by dx . We have

$$\begin{aligned}
 \hat{\alpha}_n - \alpha_0 &= \int u dF_{nh, \hat{\beta}_n}(u) - \int u dF_0(u) = \int \{F_0(u) - F_{nh, \hat{\beta}_n}(u)\} du \\
 &= \int \frac{F_0(t - \hat{\beta}'_n x) - F_{nh, \hat{\beta}_n}(t - \hat{\beta}'_n x)}{f_{T - \hat{\beta}'_n X}(t - \hat{\beta}'_n x)} dG(t, x) \\
 (S12.11) \quad &= \int \frac{F_0(t - \hat{\beta}'_n x) - F_0(t - \beta'_0 x)}{f_{T - \hat{\beta}'_n X}(t - \hat{\beta}'_n x)} dG(t, x) + \int \frac{F_0(t - \beta'_0 x) - F_{nh, \hat{\beta}_n}(t - \hat{\beta}'_n x)}{f_{T - \hat{\beta}'_n X}(t - \hat{\beta}'_n x)} dG(t, x)
 \end{aligned}$$

For the first term in the last expression we get

$$\begin{aligned}
 &\int \frac{F_0(t - \hat{\beta}'_n x) - F_0(t - \beta'_0 x)}{f_{T - \hat{\beta}'_n X}(t - \hat{\beta}'_n x)} dG(t, x) \\
 &= \int \{F_0(u) - F_0(u + x'(\hat{\beta}_n - \beta_0))\} f_{X|T - \hat{\beta}'_n X}(x|T - \hat{\beta}'_n X = u) du dx \\
 &\sim - \int x'(\hat{\beta}_n - \beta_0) f_0(u) f_{X|T - \beta'_0 X}(x|T - \beta'_0 X = u) du dx \\
 &\sim - \left\{ \int E_{\beta_0}\{X'|T - \beta'_0 X = u\} f_0(u) du \right\} (\hat{\beta}_n - \beta_0)
 \end{aligned}$$

This term, multiplied with \sqrt{n} , is asymptotically normal, with expectation zero and variance

$$\sigma_1^2 \stackrel{\text{def}}{=} a(\beta_0)' I_\epsilon(\beta_0)^{-1} a(\beta_0),$$

where $a(\beta_0)$ is the k -dimensional vector, defined by

$$a(\beta_0) = \int E_{\beta_0}\{X|T - \beta'_0 X = u\} f_0(u) du.$$

For the second term in (S12.11), we first note that,

$$\begin{aligned}
 (S12.12) \quad &F_{nh, \hat{\beta}_n}(t - \hat{\beta}'_n x) - F_0(t - \beta'_0 x) = \frac{\int \{\delta - F_0(t - \beta'_0 x)\} K_h(t - \hat{\beta}'_n x - u + \hat{\beta}'_n y) d\mathbb{P}_n(u, y, \delta)}{g_{nh, \hat{\beta}_n}(t - \hat{\beta}'_n x)}.
 \end{aligned}$$

We write (S12.12) as the sum of the integral over dP_0 and the integral over $d(\mathbb{P}_n - P_0)$ and show that the contribution of the dP_0 integral, evaluated in (S12.11) is negligible and that the contribution of the $d(\mathbb{P}_n - P_0)$ integral will yield an asymptotic normal distribution.

We have

$$\begin{aligned}
& \int \{\delta - F_0(t - \beta'_0 x)\} K_h(t - \hat{\beta}'_n x - u + \hat{\beta}'_n y) dP_0(u, y, \delta) \\
&= \int \{F_0(u - \beta'_0 y) - F_0(t - \beta'_0 x)\} K_h(t - \hat{\beta}'_n x - u + \hat{\beta}'_n y) dG(u, y) \\
&= \int \{F_0(v + (\hat{\beta}_n - \beta_0)y) - F_0(t - \beta'_0 x)\} K_h(t - \hat{\beta}'_n x - v) \\
&\quad \cdot f_{T-\hat{\beta}'_n X}(v) f_{X|T-\hat{\beta}'_n X}(y|T - \hat{\beta}'_n X = v) dv dy \\
&= f_{T-\hat{\beta}'_n X}(t - \hat{\beta}'_n x) \int \{F_0(t - \hat{\beta}'_n x + (\hat{\beta}_n - \beta_0)y) - F_0(t - \beta'_0 x)\} \\
&\quad \cdot f_{X|T-\hat{\beta}'_n X}(y|T - \hat{\beta}'_n X = t - \hat{\beta}'_n x) dy + O_p(h^2) \\
&= f_{T-\hat{\beta}'_n X}(t - \hat{\beta}'_n x) f_0(t - \beta'_0 x) (\hat{\beta}_n - \beta_0)' E\{X - x|T - \hat{\beta}'_n X = t - \hat{\beta}'_n x\} \\
&\quad + O_p(h^2) + o_p(\|\hat{\beta}_n - \beta_0\|),
\end{aligned}$$

where $\|x\|$ is the euclidean norm of the vector x . Hence we get

$$\begin{aligned}
& \int \frac{\int \{\delta - F_0(t - \beta'_0 x)\} K_h(t - \hat{\beta}'_n x - u + \hat{\beta}'_n y) dP_0(u, y, \delta)}{g_{nh, \hat{\beta}_n}(t - \hat{\beta}'_n x) f_{T-\hat{\beta}'_n X}(t - \hat{\beta}'_n x)} dG(t, x) \\
&= (\hat{\beta}_n - \beta_0)' \int \frac{f_0(t - \beta'_0 x) E\{X - x|T - \beta'_0 X = t - \beta'_0 x\}}{g_{nh}(t - \hat{\beta}'_n x)} dG(t, x) + O_p(h^2) + o_p(\|\hat{\beta}_n - \beta_0\|) \\
&= (\hat{\beta}_n - \beta_0)' \int f_0(v) E\{X - x|T - \beta'_0 X = v\} f_{X|T-\beta'_0 X}(x|T - \beta'_0 X = v) dx dv \\
&\quad + O_p(h^2) + o_p(\|\hat{\beta}_n - \beta_0\|) \\
&= O_p(h^2) + o_p(\|\hat{\beta}_n - \beta_0\|),
\end{aligned}$$

which is $o_p(n^{-1/2})$ if $h \ll n^{-1/4}$.

Finally,

$$\begin{aligned}
& \sqrt{n} \int \frac{\int \{\delta - F_0(t - \beta'_0 x)\} K_h(t - \hat{\beta}'_n x - u + \hat{\beta}'_n y) d(\mathbb{P}_n - P_0)(u, y, \delta)}{g_{nh, \hat{\beta}_n}(t - \hat{\beta}'_n x) f_{T-\hat{\beta}'_n X}(t - \hat{\beta}'_n x)} dG(t, x) \\
&= \sqrt{n} \iint \frac{\{\delta - F_0(t - \beta'_0 x)\} K_h(t - \hat{\beta}'_n x - u + \hat{\beta}'_n y)}{g_{nh, \hat{\beta}_n}(t - \hat{\beta}'_n x) f_{T-\hat{\beta}'_n X}(t - \hat{\beta}'_n x)} dG(t, x) d(\mathbb{P}_n - P_0)(u, y, \delta) \\
&= \sqrt{n} \int \frac{\{\delta - F_0(u - \beta'_0 y)\}}{f_{T-\beta'_0 X}(u - \beta'_0 y)} d(\mathbb{P}_n - P_0)(u, y, \delta) + O_p(h^2) + O_p(\|\hat{\beta}_n - \beta_0\|)
\end{aligned}$$

is asymptotically normal, with expectation zero and variance

$$(S12.13) \quad \int \frac{F_0(v)\{1 - F_0(v)\}}{f_{T-\beta'_0 X}(v)} dv,$$

if $h \ll n^{-1/4}$.

Both terms in the representation on the right of (S12.11) are, apart from a negligible contribution, sums of independent variables with expectation zero. By Theorem 4.5 we have

$$\begin{aligned} & \sqrt{n}(\hat{\beta}_n - \beta_0) \\ &= n^{-1/2} I_\epsilon(\beta_0)^{-1} \sum_{i=1}^n f_0(T_i - \beta_0 X_i) \{E_{\beta_0}(X_i | T_i - \beta'_0 X_i) - X_i\} \\ & \quad \cdot \frac{\Delta_i - F_0(T_i - \beta'_0 X_i)}{F_0(T_i - \beta'_0 X_i) \{1 - F_0(T_i - \beta'_0 X_i)\}} 1_{[\epsilon, 1-\epsilon]} \{F_0(T_i - \beta'_0 X_i)\} + o_p(1). \end{aligned}$$

and the second term of (S12.11) has the representation

$$n^{-1/2} \sum_{i=1}^n \frac{\Delta_i - F_0(T_i - \beta'_0 X_i)}{f_{T-\beta'_0 X}(T_i - \beta'_0 X_i)}.$$

By the independence of the summands with indices $i \neq j$, the only contribution to the covariance of the two terms in the representation can come from summands with the same index. But,

$$\begin{aligned} & E_{\beta_0} \left\{ \frac{f_0(T_i - \beta'_0 X_i) \{E_{\beta_0}(X_i | T_i - \beta'_0 X_i) - X_i\} \{\Delta_i - F_0(T_i - \beta'_0 X_i)\}^2}{F_0(T_i - \beta'_0 X_i) \{1 - F_0(T_i - \beta'_0 X_i)\} f_{T-\beta'_0 X}(T_i - \beta'_0 X_i)} 1_{[\epsilon, 1-\epsilon]} \{F_0(T_i - \beta'_0 X_i)\} \right\}. \\ &= \int_{F_0(u - \beta'_0 y) \in [\epsilon, 1-\epsilon]} \frac{f_0(u - \beta'_0 y) \{E_{\beta_0}(X | T - \beta'_0 X = u - \beta'_0 y) - y\} \{\delta - F_0(u - \beta'_0 y)\}^2}{F_0(u - \beta'_0 y) \{1 - F_0(u - \beta'_0 y)\} f_{T-\beta'_0 X}(u - \beta'_0 y)} dP_0(u, y, \delta) \\ &= \iint_{F_0(v) \in [\epsilon, 1-\epsilon]} \frac{f_0(v) \{E_{\beta_0}(X | T - \beta'_0 X = v) - y\} F_0(v) \{1 - F_0(v)\}}{F_0(v) \{1 - F_0(v)\}} f_{X|T-\beta'_0 X}(y|v) dv dy \\ &= \int_{F_0(v) \in [\epsilon, 1-\epsilon]} \int \{E_{\beta_0}(X | T - \beta'_0 X = v) - y\} f_{X|T-\beta'_0 X}(y|v) dy \frac{f_0(v) F_0(v) \{1 - F_0(v)\}}{F_0(v) \{1 - F_0(v)\}} dv \\ &= 0 \end{aligned}$$

So the covariance is zero and Theorem 5.1 follows. \square

S13. Acknowledgements. The authors want to thank Richard Nickl for useful comments and his reference to relevant theory in the book [11] and Fadoua Balabdaoui for incisive questions and references to relevant literature. We also thank the associate editor and two referees for their valuable remarks. The research of the second author was supported by the Research Foundation Flanders (FWO) [grant number 11W7315N]. Support from the IAP Research Network P7/06 of the Belgian State (Belgian Science Policy) is gratefully acknowledged.

References.

- [1] BIRMAN, M. Š. and SOLOMJAK, M. Z. (1967). Piecewise polynomial approximations of functions of classes W_p^α . *Mat. Sb. (N.S.)* **73** (115) 331–355. [MR0217487 \(36 #576\)](#)
- [2] CHEN, X., LINTON, O. and VAN KEILEGOM, I. (2003). Estimation of semiparametric models when the criterion function is not smooth. *Econometrica* **71** 1591–1608.
- [3] COSSLETT, S. R. (1983). Distribution-free maximum likelihood estimator of the binary choice model. *Econometrica* **51** 765–782. [MR712369 \(85a:62174\)](#)
- [4] COSSLETT, S. R. (1987). Efficiency bounds for distribution-free estimators of the binary choice and the censored regression models. *Econometrica* **55** 559–585. [MR890854 \(88e:62087\)](#)
- [5] COSSLETT, S. R. (2007). Efficient estimation of semiparametric models by smoothed maximum likelihood. *Internat. Econom. Rev.* **48** 1245–1272. [MR2375624](#)
- [6] DING, Y. and NAN, B. (2011). A sieve M -theorem for bundled parameters in semiparametric models, with application to the efficient estimation in a linear model for censored data. *Ann. Statist.* **39** 3032–3061. [MR3012400](#)
- [7] DOMINITZ, J. and SHERMAN, R. P. (2005). Some convergence theory for iterative estimation procedures with an application to semiparametric estimation. *Econometric Theory* **21** 838–863. [MR2189497 \(2006h:60039\)](#)
- [8] FINKELSTEIN, D. M. (1986). A proportional hazards model for interval-censored failure time data. *Biometrics* **42** 845–854. [MR872963 \(87m:62286\)](#)
- [9] FINKELSTEIN, D. M. and WOLFE, R. A. A semiparametric model for regression analysis of interval-censored failure time data. *Biometrics* **41** 933–945.
- [10] GESKUS, R. B. and GROENEBOOM, P. (1997). Asymptotically optimal estimation of smooth functionals for interval censoring. II. *Statist. Neerlandica* **51** 201–219. [MR1466426 \(99d:62015\)](#)
- [11] GINÉ, E. and NICKL, R. (2015). *Mathematical Foundations of Infinite-Dimensional Statistical Models*. Cambridge Series in Statistical and Probabilistic Mathematics.
- [12] GROENEBOOM, P. and HENDRICKX, K. (2016). Supplement to “Current status linear regression”.
- [13] GROENEBOOM, P., JONGBLOED, G. and WITTE, B. I. (2010). Maximum smoothed likelihood estimation and smoothed maximum likelihood estimation in the current status model. *Ann. Statist.* **38** 352–387.
- [14] GROENEBOOM, P. and JONGBLOED, G. (2014). *Nonparametric Estimation under Shape Constraints*. Cambridge Univ. Press, Cambridge.
- [15] GROENEBOOM, P. and WELLNER, J. A. (1992). *Information bounds and nonparametric maximum likelihood estimation*. *DMV Seminar* **19**. Birkhäuser Verlag, Basel. [MR1180321 \(94k:62056\)](#)
- [16] HAN, A. K. (1987). Non-parametric analysis of a generalized regression model: the maximum rank correlation estimator. *Journal of Econometrics* **35** 303–316.
- [17] HÄRDLE, W., HALL, P. and ICHIMURA, H. (1993). Optimal smoothing in single-index models. *Ann. Statist.* **21** 157–178. [MR1212171 \(94d:62134\)](#)

- [18] HOEFFDING, W. (1940). Masstabinvariante Korrelationtheorie. *Schriften Math. Inst. Univ. Berlin* **5** 181–233. Translated in: The Collected Works of Wassily Hoeffding, N.I. Fisher and P.K. Sen (eds.), Springer Verlag, New York 1994.
- [19] HUANG, J. (1995). Maximum likelihood estimation for proportional odds regression model with current status data. In *Analysis of censored data (Pune, 1994/1995). IMS Lecture Notes Monogr. Ser.* **27** 129–145. Inst. Math. Statist., Hayward, CA. [MR1483344](#)
- [20] HUANG, J. and WELLNER, J. (1993). Regression models with interval censoring. *Proceedings of the Kolmogorov Seminar, Euler Mathematics Institute. St. Petersburg, Russia.*
- [21] KLEIN, R. W. and SPADY, R. H. (1993). An efficient semiparametric estimator for binary response models. *Econometrica* **61** 387–421. [MR1209737 \(93k:62214\)](#)
- [22] KOSOROK, M. R. (2008). Bootstrapping the Grenander estimator. In *Beyond parametrics in interdisciplinary research: Festschrift in honor of Professor Pranab K. Sen. Inst. Math. Stat. Collect.* **1** 282–292. Inst. Math. Statist., Beachwood, OH.
- [23] LI, G. and ZHANG, C.-H. (1998). Linear regression with interval censored data. *Ann. Statist.* **26** 1306–1327. [MR1647661 \(2000b:62094\)](#)
- [24] MURPHY, S. A., VAN DER VAART, A. W. and WELLNER, J. A. (1999). Current status regression. *Math. Methods Statist.* **8** 407–425. [MR1735473 \(2001a:62126\)](#)
- [25] POLLARD, D. (1984). *Convergence of stochastic processes. Springer Series in Statistics.* Springer-Verlag, New York. [MR762984 \(86i:60074\)](#)
- [26] RABINOWITZ, D., TSIATIS, A. and ARAGON, J. (1995). Regression with interval censored data. *Biometrika* **82** 501–513.
- [27] ROSSINI, A. and TSIATIS, A. A. (1996). A semiparametric proportional odds regression model for the analysis of current status data. *Journal of American Statistical Association* **91** 713–721.
- [28] SEN, B., BANERJEE, M. and WOODROOFE, M. B. (2010). Inconsistency of bootstrap: the Grenander estimator. *Ann. Statist.* **38** 1953–1977. [MR2676880 \(2011f:62046\)](#)
- [29] SEN, B. and XU, G. (2015). Model based bootstrap methods for interval censored data. *Comput. Statist. Data Anal.* **81** 121–129. [MR3257405](#)
- [30] SHEN, X. (2000). Linear regression with current status data. *J. Amer. Statist. Assoc.* **95** 842–852. [MR1804443 \(2001k:62118\)](#)
- [31] SHERMAN, R. P. (1993). The limiting distribution of the maximum rank correlation estimator. *Econometrica* **61** 123–137. [MR1201705](#)
- [32] SHIBOSKI, S. C. and JEWELL, N. P. (1992). Statistical Analysis of the Time Dependence of HIV Infectivity Based on Partner Study Data. *Journal of the American Statistical Association* **87** 360–372.
- [33] VAN DER VAART, A. W. (1998). *Asymptotic statistics. Cambridge Series in Statistical and Probabilistic Mathematics* **3**. Cambridge University Press, Cambridge. [MR1652247 \(2000c:62003\)](#)

DELFT UNIVERSITY OF TECHNOLOGY, MEKELWEG 4, 2628 CD DELFT, THE NETHERLANDS.

E-MAIL: P.Groeneboom@tudelft.nl

HASSELT UNIVERSITY, I-BioSTAT, AGORALAAN, B3590 DIEPENBEEK, BELGIUM.

E-MAIL: kim.hendrickx@uhasselt.be